

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2002, 09:22:42 ; Search time 41.23 Seconds
(without alignments)
4212.634 Million cell updates/sec

Title: US-09-767-215-2

Perfect score: 5149

Sequence: 1 MGELCRDRSALTALDEETLW.....VRQAIADQKKVVWTEQSPR 1004

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19.*

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-----------|--------------------|
| 1 | 5149 | 100.0 | 1004 | Q9BXL6 | Q9bxl6 homo sapien |
| 2 | 2484.5 | 48.3 | 662 | 11 Q99KF0 | Q99kf0 mus musculu |
| 3 | 1910 | 37.1 | 434 | 4 Q9BVB5 | Q9bvb5 homo sapien |
| 4 | 1239.5 | 24.1 | 1147 | 4 Q9BXL7 | Q9bxl7 homo sapien |
| 5 | 887.5 | 17.2 | 1032 | 4 Q9BWT7 | Q9bwt7 homo sapien |
| 6 | 512.5 | 10.0 | 536 | 11 Q9EPY0 | Q9epy0 rattus norv |
| 7 | 493 | 9.6 | 536 | 4 Q9H257 | Q9h257 homo sapien |
| 8 | 396.5 | 7.7 | 366 | 4 Q9H854 | Q9h854 homo sapien |
| 9 | 351.5 | 6.8 | 1695 | 5 Q9BKL2 | Q9bkl2 hydra atten |
| 10 | 315 | 6.1 | 332 | 4 Q9UGR6 | Q9ugr6 homo sapien |
| 11 | 307.5 | 6.0 | 331 | 4 Q9UGR5 | Q9ugr5 homo sapien |
| 12 | 290.5 | 5.6 | 1769 | 6 Q97F58 | Q97f58 canis famil |
| 13 | 290 | 5.6 | 1313 | 4 Q75033 | Q75033 homo sapien |
| 14 | 282 | 5.5 | 974 | 5 Q960N4 | Q960n4 drosophila |
| 15 | 282 | 5.5 | 1367 | 5 Q94880 | Q94880 drosophila |
| 16 | 282 | 5.5 | 1445 | 5 Q9VHK4 | Q9vhk4 drosophila |

| | | | | | |
|----|-------|-----|------|-----------|---------------------|
| 17 | 262 | 5.1 | 2101 | 4 Q14981 | Q14981 homo sapien |
| 18 | 259 | 5.0 | 813 | 11 P70625 | P70625 rattus norv |
| 19 | 258 | 5.0 | 1916 | 5 Q9VKG8 | Q9vkg8 drosophila |
| 20 | 254 | 4.9 | 304 | 11 Q921G9 | Q921g9 mus musculu |
| 21 | 253 | 4.9 | 1257 | 5 Q9NF04 | Q9nf04 caenorhabdi |
| 22 | 252 | 4.9 | 2115 | 4 Q14980 | Q14980 homo sapien |
| 23 | 248.5 | 4.8 | 2442 | 4 Q14812 | Q14812 homo sapien |
| 24 | 246 | 4.8 | 2611 | 11 Q912D8 | Q91zu8 mus musculu |
| 25 | 245 | 4.8 | 746 | 5 Q25561 | Q25561 naegleria f |
| 26 | 244.5 | 4.7 | 2442 | 4 Q9H450 | Q9h450 homo sapien |
| 27 | 243.5 | 4.7 | 2168 | 11 Q61043 | Q61043 mus musculu |
| 28 | 242.5 | 4.7 | 3259 | 4 Q14789 | Q14789 homo sapien |
| 29 | 242 | 4.7 | 1163 | 13 Q9YHV2 | Q9yvhv2 gallus gall |
| 30 | 241.5 | 4.7 | 2442 | 4 Q60588 | Q60588 homo sapien |
| 31 | 236 | 4.6 | 1200 | 11 Q921B9 | Q921b9 mus musculu |
| 32 | 235.5 | 4.6 | 1692 | 4 Q9P216 | Q9p216 homo sapien |
| 33 | 235.5 | 4.6 | 1889 | 4 Q9H430 | Q9h430 homo sapien |
| 34 | 235 | 4.6 | 612 | 4 Q96JV2 | Q96jv2 homo sapien |
| 35 | 235 | 4.6 | 2029 | 4 Q9C014 | Q9c014 homo sapien |
| 36 | 235 | 4.6 | 2069 | 4 Q9C013 | Q9c013 homo sapien |
| 37 | 235 | 4.6 | 2073 | 4 Q9C012 | Q9c012 homo sapien |
| 38 | 233.5 | 4.5 | 1456 | 5 Q9V587 | Q9v587 drosophila |
| 39 | 230 | 4.5 | 1281 | 4 Q9Y4E3 | Q9y4e3 homo sapien |
| 40 | 229.5 | 4.5 | 2041 | 4 Q9HAT7 | Q9hat7 homo sapien |
| 41 | 229.5 | 4.5 | 2047 | 4 Q9UH61 | Q9uh61 homo sapien |
| 42 | 228 | 4.4 | 674 | 4 Q9UE73 | Q9ue73 homo sapien |
| 43 | 227.5 | 4.4 | 2473 | 11 Q9Q284 | Q9qz84 mus musculu |
| 44 | 226.5 | 4.4 | 1270 | 4 Q96JN2 | Q96jn2 homo sapien |
| 45 | 223.5 | 4.3 | 2363 | 4 Q99968 | Q99968 homo sapien |

ALIGNMENTS

RESULT 1

Q9BXL6 ID Q9BXL6 PRELIMINARY; PRT; 1004 AA.
AC Q9BXL6;
DT 01-JUN-2001 (TREMREL. 17, Created)
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)
DT 01-OCT-2001 (TREMREL. 18, Last annotation update)
DE CASPASE RECRUITMENT DOMAIN PROTEIN 14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21192234; PubMed=11278692;
RA Bertin J., Wang L., Guo Y., Jacobson M.D., Poyet J.-L.,
RA Srinivasula S.M., Merriam S., Distefano P.S., Alnemri E.S.;
RT "CARD11 and CARD14 Are Novel Caspase Recruitment Domain
(CARD)/Membrane-associated Guanylate Kinase (MAGUK) Family Members
that Interact with Bcl10 and Activate NF-kappaB.";
RL J. Biol. Chem. 276:11877-11882(2001).
DR EMBL; AF322642; AAG53403.1; -
DR InterPro; IPR000619; Guanylate_kin.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM00072; GUKC; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 1.
SQ SEQUENCE 1004 AA; 113299 MW; 7469B8B56BE06073 CRC64;

Query Match 100.0%; Score 5149; DB 4; Length 1004;
Best Local Similarity 100.0%; Pred. No. 1.5e-295;
Matches 1004; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGELCRDRSALTALDEETLWEMMESHRRIVRCICPSRLTPYLROAKVLCOLDEEVLHS 60

DB 1 MGELCRDRSALTALDEETLWEMMESHRRIVRCICPSRLTPYLROAKVLCOLDEEVLHS 60

QY 61 PRLTNSAMRAGHLLDLKTRGKNGAIAFLSKFHNPDVYTLVTGLQPDVDFSNFSGLME 120
 DB 61 PRLTNSAMRAGHLLDLKTRGKNGAIAFLSKFHNPDVYTLVTGLQPDVDFSNFSGLME 120
 QY 121 TSKLTECLAGAGISQBELNQKQEVLLRRCQQLQEHGLAETRAEGIHQLEADHSRM 180
 DB 121 TSKLTECLAGAGISQBELNQKQEVLLRRCQQLQEHGLAETRAEGIHQLEADHSRM 180
 QY 181 KREVSASFHVLRKDEMLSLSHYSNALQEKELASRCSLQBELYLKQELQORANMVS 240
 DB 181 KREVSASFHVLRKDEMLSLSHYSNALQEKELASRCSLQBELYLKQELQORANMVS 240
 QY 241 SCELEQEQSLRTASQESGDEELNRLKEENKRLSLTFLSAEKDILEQSLDEARGSRQE 300
 DB 241 SCELEQEQSLRTASQESGDEELNRLKEENKRLSLTFLSAEKDILEQSLDEARGSRQE 300
 QY 301 LVERIHSRERAVAAERQRYWEKEQTLQFQKSKMACOLYREKVNALQAOVCELOKE 360
 DB 301 LVERIHSRERAVAAERQRYWEKEQTLQFQKSKMACOLYREKVNALQAOVCELOKE 360
 QY 361 RDOAYSARDSAQEISQSLVEKSLRROVFEITDQVCELRQLQLOAEPGVVLKQEART 420
 DB 361 RDOAYSARDSAQEISQSLVEKSLRROVFEITDQVCELRQLQLOAEPGVVLKQEART 420
 QY 421 REPCPREKQRLVRMHATCPRDDSDCSLVSTESQSLSDLSATSSRELVDSEFRSSPPAPS 480
 DB 421 REPCPREKQRLVRMHATCPRDDSDCSLVSTESQSLSDLSATSSRELVDSEFRSSPPAPS 480
 QY 481 QOSLYKRVAEDEGEPWFSFSCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSL 540
 DB 481 QOSLYKRVAEDEGEPWFSFSCLEIPEGDPGALPGAKAGDPHLDYELLDTADLPQLESSL 540
 QY 541 QVSPGRLDVSSEGLMRRPARRILSQVTMLAFQGDALLEQISVIGNLTGFIHRTVP 600
 DB 541 QVSPGRLDVSSEGLMRRPARRILSQVTMLAFQGDALLEQISVIGNLTGFIHRTVP 600
 QY 601 GSAADQALRPGTQIVMVDYEAEPFLKAVLEDTTLEAVGLLRVDGFCCLSVKYNVDG 660
 DB 601 GSAADQALRPGTQIVMVDYEAEPFLKAVLEDTTLEAVGLLRVDGFCCLSVKYNVDG 660
 QY 661 YKRLQDLQAKVATSGDSFYIRVNLAEGRAKGELOVHCNEVLHVTDVTFQGGCGWHAHR 720
 DB 661 YKRLQDLQAKVATSGDSFYIRVNLAEGRAKGELOVHCNEVLHVTDVTFQGGCGWHAHR 720
 QY 721 VNSTYMTKDTAAHGTPINYRAQOQLIALIODMTQOCTVTRKPSGGQPKLVIRVSMDBKAK 780
 DB 721 VNSTYMTKDTAAHGTPINYRAQOQLIALIODMTQOCTVTRKPSGGQPKLVIRVSMDBKAK 780
 QY 781 ASPLRLSFRDQCLDPSRMESGSCFVAESCLTLVPYTLVMPHRAERPRVLLVPRAVGKI 840
 DB 781 ASPLRLSFRDQCLDPSRMESGSCFVAESCLTLVPYTLVMPHRAERPRVLLVPRAVGKI 840
 QY 841 LSEKLCLOGKFKCLAEVLSQEEYEAWSQSGDIIQEGEVSGGRCWTVRHAVESLMEXNTH 900
 DB 841 LSEKLCLOGKFKCLAEVLSQEEYEAWSQSGDIIQEGEVSGGRCWTVRHAVESLMEXNTH 900
 QY 901 ALLDQVLDVSCTLRHMDIFPIVHVSVNEKMAKKLKGLORLGTSEBOLLEAARQEGDL 960
 DB 901 ALLDQVLDVSCTLRHMDIFPIVHVSVNEKMAKKLKGLORLGTSEBOLLEAARQEGDL 960
 QY 961 DRAPCLYSSAPDGSDDLGLLSCVROAIADQEKVWVTEQSPR 1004
 DB 961 DRAPCLYSSAPDGSDDLGLLSCVROAIADQEKVWVTEQSPR 1004

RESULT 2
 Q99KF0 PRELIMINARY: PRT: 662 AA.
 AC Q99KF0;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE HYPOTHETICAL 75.4 KDA PROTEIN (FRAGMENT).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC004692; AAH04692.1; -.
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00595; PDZ; 1.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 662 AA; 75422 MW; AADFE37A7A04513D CRC64;

Query Match 48.3%; Score 2484.5; DB 11; Length 662;
 Best Local Similarity 75.1%; Pred. No. 1.3e-138;
 Matches 505; Conservative 63; Mismatches 89; Indels 15; Gaps 5;

QY 82 KNGAIAFLSKFHNPDVYTLVTGLQPDVDFSNFSGLMETSKLTCIAGAGISQBELNQ 141
 DB 1 KNGAIAFLSKFHNPDVYTLVTGLQSDIDFSTFSGLMETSKLTCIAGAGISQBELAQ 60
 QY 142 EKQKEVLLRRCQQLQEHGLAETRAEGHLQLEADHSRMRKREYSAHFHEVLRKDEMLSL 201
 DB 61 EKAQKEVLLRRCQQLKRLGLAEHAEGRLQLEVDHSRMRKREYSTHFHEVLRKDEMLNL 120
 QY 202 SLHYSNALQEKELASRCSLQBELYLKQELQORANMVSSCELELQBOQSLRTADQESGD 261
 DB 121 SLHYSNALREKELAATRCHSLQBELYLKQELQORASLVSSCRESRESRLKMASNLEPQG 180
 QY 262 EELNRLKEENKRLSLTFLSAEKDILEQSLDEARGSRQELVERIHSRERAVAAERQEQ 321
 DB 181 EELNRLKEENKRLSRMTFSLVEKDILEQSLDEARGSRQELVDRHLSRERAVAAERQEQ 240
 QY 322 YWEEKOTLLOFKSKMACOLYREKVNALQAOVCELRQDQAYSARDSAQEISQSLVE 381
 DB 241 YWEEKOTLLOFKRKTQVDCELYKEKMTLQGVLAELQKQERQAYTARDRAQMEISQRLVE 300
 QY 382 KDSLRROVFEITDQVCELRQLQLOAEPGVVLKQEAETREPCPREKQRLVRMHATCPRD 441
 DB 301 KDLRRRVFEITDQVCELRQLRLQAEAPGPKQAEAGARELCRLRGKQRLVRMHAVCPD 360
 QY 442 DSDCLSVSTESQSLSDLSATSSRELVDSEFRSSPPAPSQQSLYKRVAEDEGEPWFSFS 501
 DB 361 DSDCLSVSTESRLWDLNSTSSREQMDSFRSSPPMPSSQSLYKRVAEDEGEPWFSFS 417
 QY 502 CLEIPEGDQALPGAKAG--DPHLDYELLDTADLPQLESSLPQVSPGRLDVSSEGLMRR 559
 DB 418 -LSFPEVLEMRLOQATVDDTDTDLFEFEMIDGADLSQTESDLSQG--SSRSLANVSESSVPR 475
 QY 560 RPARRILSQVTMLAFQGDALLEQISVIGNLTGFIHRTVPGSAAQOMALRPGTQIVMVD 619
 DB 476 RPARRILSQVTMLAFQGDALLEQISVIGNLTGFIHRTVPGSAAQOMALRPGTQIVMVD 535
 QY 620 YEASEPLFKAVLEDTTLEAVGLLRVDGFCCLSVKYNVDGKRLQDLQAKVATSGDSF 679
 DB 536 YKPTKPSLRATLENTLEQAVGLLRVNGSCYLSVKINTEGYKNLIQDLQAKVATSGDSF 595
 QY 680 YIRVNLAEGRAKGELOVHCNEVLHVTDVTFQGGCGWHAHRVNSYTHKDTAAHGTINYS 739
 DB 596 YIRVNLAEGRAKGELOVHCNEVLHVTDVTFQGGCGWHAHRVNSYTHKDTAAHGTINYS 654
 QY 740 RAQOQLIALIOD 751
 DB 655 H-----LLED 659

RESULT 3

```

Q9BVB5
ID Q9BVB5 PRELIMINARY; PRT; 434 AA.
AC Q9BVB5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 48.5 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CERVIX, CARCINOMA;
RA Strausberg R.;
RL EMBL; BC001326; AAH01326.1; -.
RW Hypothetical protein.
SQ SEQUENCE 434 AA; 48504 MW; D351699781B6D6EB CRC64;

Query Match 37.1%; Score 1910; DB 4; Length 434;
Best Local Similarity 100.0%; Pred. No. 6e-105;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 238 MVSSCELELEQEQSLRTASDOESGDEELNRLKEENKRLSLTFLSLAEKDIILEQSLDEARGS 297
DB 1 MVSSCELELEQEQSLRTASDOESGDEELNRLKEENKRLSLTFLSLAEKDIILEQSLDEARGS 60

QY 298 RQELVERIHSLRRAVAERQROYWEEKQTLLQFQKSKMACOLYREKYNALQAQVCEL 357
DB 61 RQELVERIHSLRRAVAERQROYWEEKQTLLQFQKSKMACOLYREKYNALQAQVCEL 120

QY 358 OKERDAYASARDSAQREISQSLVEKDSLRQVFEITDQVCELTQRLQLOAEPGVKLQE 417
DB 121 OKERDAYASARDSAQREISQSLVEKDSLRQVFEITDQVCELTQRLQLOAEPGVKLQE 180

QY 418 ARTREPCPREKQRLVRMHAICPRDDSDCLSVSTESQSLSDLSATSSRELVDVDFRSSPA 477
DB 181 ARTREPCPREKQRLVRMHAICPRDDSDCLSVSTESQSLSDLSATSSRELVDVDFRSSPA 240

QY 478 PPSQSLYKRVAEDEGPEPWSFSCLEIPEGDPGALPGAKAGPHLDYELLDADLPQLE 537
DB 241 PPSQSLYKRVAEDEGPEPWSFSCLEIPEGDPGALPGAKAGPHLDYELLDADLPQLE 300

QY 538 SSIQPVSPGRLDVSESGVLMRRPARRILSQVTMLAFQGDALLEQISVIGNLTGFIHR 597
DB 301 SSIQPVSPGRLDVSESGVLMRRPARRILSQVTMLAFQGDALLEQISVIGNLTGFIHR 360

QY 598 VTPGSAADQALRPGTQIVMV 618
DB 361 VTPGSAADQALRPGTQIVMV 381

RESULT 4
Q9BXL7
ID Q9BXL7 PRELIMINARY; PRT; 1147 AA.
AC Q9BXL7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE CASPASE RECRUITMENT DOMAIN PROTEIN 11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21192234; PubMed=11278692;
RA Bertin J., Wang L., Guo Y., Jacobson M.D., Poyet J.-L.,
RA Srinivasula S.M., Merriam S., DiStefano P.S., Alnemri E.S.;
RT "CARD11 and CARD14 Are Novel Caspase Recruitment Domain
(CARD)/Membrane-associated Guanylate Kinase (MAGUK) Family Members

```

```

RT that Interact with BCL10 and Activate NF-kappaB.";
RL J. Biol. Chem. 276:11877-11882(2001).
DR EMBL; AF322641; AGS3402.1; -.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
SQ SEQUENCE 1147 AA; 132641 MW; 913A4B015D2B36CC CRC64;

Query Match 24.1%; Score 1239.5; DB 4; Length 1147;
Best Local Similarity 30.5%; Pred. No. 7.9e-65;
Matches 354; Conservative 204; Mismatches 402; Indels 201; Gaps 33;

QY 15 DEETLWEMESHRRIRVRCICPSRLTPYLRQAKVLCQDEEVLHSPRLTNSAMRAGHLL 74
DB 11 EEDALWENVECNHRLSRYINPAKLTPLYRQCKVIDEODEVLNAPMLPSKINRAGRL 70

QY 75 DLLKTRKNGAIAIPLESKLFHNPVYITVGLQPDVDFSNFSGLMETSKLTCELAGAIGS 134
DB 71 DILHTKGQGVVVFLESLEFYYPELYKLTGKPTRRFTIVVEEGHGLTHFLMNEVIK 130

QY 135 LOELNQEKGQEVLLRRCOLO-EHLGLAETRAEGLHLEADHSRMKREVSAAHFVLR 193
DB 131 LQQMKAKDLQRCCLARLQLEDEKKOMTLTRVE-LTFOERYKKMKREDSYNDELVK 189

QY 194 LKDEMLSLSHYSNALQEKELAAASRCRSLQBELLYLKQELQRANWSSCELELEQESLRT 253
DB 190 VKDNYNLAMRYAQLSEKNVAVMRSRDQLQLEIDLQHLRNLK--MEBECKLE-RNQSLK 246

QY 254 ASDQES--GDELNRLKEENKRLSLTFLS-----AKDILQESLDEAR 295
DB 247 KNDIENPKKEQVLELEERENMLTKNQELQSIITQAGKRSLPDSDKATILDEHDKAL 306

QY 296 GSQELVERIHSLRRAVAERQROYWEEKQTLLQFQKSKMACOLYREKYNALQAQVC 355
DB 307 EDQELVRYINLQEEARQAEELKDYEEKDELEKSTLGKDCCKMTHKRNVTVMLOLE 366

QY 356 ELQERDAYASARDSAQREISQSLVEKDSLRQVFEITDQ-----VCELRT 401
DB 367 EVERERDQAFHSRDEAQTQVSCLEIKDKYRKQIRELEENKDEMREMVRREACIVNLES 426

QY 402 QLRLQLE-----PPGVKLQ-----EARTREPCPREK----- 428
DB 427 KLRLSDSNLNDQSLRPNLPTVITISQDFGASPRNGEADDSSTSESPESKYFLPY 486

QY 429 ---QRLVRMHAICPRDDSDCLSVSTESQSL-----SDLSATS--SRELVDVDFRSSP 476
DB 487 HPORRNKLGILQORAKSPISLARTSDFQAKGHEEGTDAASPCGSLPTINSFTKMQP 546

QY 477 -----APPSQSLYKRVAEDEGPEPWSFSCLEIPEGDPGALPGAKA-GDPHL 523
DB 547 PRSRSSIMSTAEPPGNDSTVRYKEDAPHR-----STVE-EDNDSGGFDALDLDSDSHE 600

QY 524 DY-----ELLDADLPQLESLOPV-----PGRLDVSESGVLMRRPARR 564
DB 601 RYSEFGPSSIHSSSSSHQSEGDAYDLEQVNLNMFKFSLEPRFSPSVTSVGHVRGPGPS-- 658

QY 565 ILSQVTMLAFQGDALLEQISVIGNLTGFIHRVTPGSAADQALRPGTQIVMVDYEA 624
DB 659 ---VQHTTLNGDSLTQTLTLLGNARGSFVHSVKPSLAEGKAGLREGHQJLLLEGCTIRG 714

QY 625 PLFKAVILEDTTLEBAVGLLRVRDGCCLSVKVNITDGKRYRLQDLQLEAKVATSGDSFYIRVN 684
DB 715 ERQSVPLDTCTKEEAHTIQRCSGPTVLRHYKVNHEGYRKLKVKMEDGLITSGDSFYIRLN 774

QY 685 LAMEGRAGK-ELQVHCNEVLHVDTMFCGCCGCHAHRVNSVTMKDTHAAGTIPNYSRAQ 743
DB 775 LNISSQDACTMSLKDDVVHVRDTMYQDRHENPCARVDPFTDHDLDN-GTIPYSRAQ 833

QY 744 QLIALIOMTQO-----CTVTRKP-----SSGGPQ-----KLV 771
DB 834 LLVVKLQRLMHRGSEEVGDGTHHTLRALNLTQPEALSTSDPRVPSRLSRASFLGOLL 893

```

```

QY 772 RIVSMKAK-----ASPL-----RLSFD-----RGOLDPSRMGSSSTCF 805
Db 894 QFVSRSENKYMNSNERNVRIISGSPGLSARSLDQATKLLTEKQBELDPESELGNK--- 950
QY 806 WAESCLTLVPTLVWHPRPAPRPVLLVPRAGVKILSEKLCLLQG---FKKCLAEYLSQE 862
Db 951 -----LSLIPLSVLRYACERRRPVLTPTVLAKTLVORLLNSGGAMEFTICKSDIVTRD 1005
QY 863 EYEAWSGDII--QEVESSGGRCWVTRHVESLMEKNTHALLDVQDSVCTLHRMDIFP 920
Db 1006 EFLRQKTETIYSREKNPNAFEC-IAPANIEAAVAAKNKHCLEAGIGCTTRDLIKSNIYP 1064
QY 921 IVIHVSNEKMAKLLKGLQRLGTSSEOLLEAARQEGDLDRAPCLYSLSLAPDQWSDLOG 980
Db 1065 IVLPIRCEKNIKRKLRLRPET-EEFLRVCRKLEKEALPCLYATVPEPDMWGSVEE 1123
QY 981 LLSVQRAIADEQKVVWTEQ 1001
Db 1124 LLRVVKDIGEQRKTIWDE 1144

RESULT 5
Q9BWT7 PRELIMINARY; PRT; 1032 AA.
ID Q9BWT7
AC Q9BWT7
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CASPASE RECRUITMENT DOMAIN PROTEIN 10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21292987; PubMed=11259443;
RA Wang L., Huang W.J., Ke X., Poyet J.L., Manji G.A.,
RA Merriam S., Glucksmann M.A., DiStefano P.S., Alnemri E.S., Bertin J.;
RT "CARD10 Is a Novel Caspase Recruitment Domain/Membrane-associated
RT Guanylate Kinase Family Member That Interacts with BCL10 and Activates
RT NF-kappa B.";
RL J. Biol. Chem. 276:21405-21409(2001).
DR EMBL; AY028896; AAK26165.1; -.
SQ SEQUENCE 1032 AA; 115946 MW; 8377319AB82A0949 CRC64;

Query Match 17.2%; Score 887.5; DB 4; Length 1032;
Best Local Similarity 29.2%; Pred. No. 4e-44;
Matches 317; Conservative 161; Mismatches 419; Indels 189; Gaps 31;

QY 15 DEETLWEMESHRRIVRICPSRLTPYLQAKVLCQDDEEVLHSPRLTNSAMRAGHLL 74
Db 23 EEDALWERIEGVHRRLARALNPAKLTPLYLRQCRVIDEQDEEVLSTYRFPCKRVNTRGLM 82
QY 75 DILKTRGKCATAFLESKFHPNDVYTLVGLQPDVDFSNFSLMETSCLKTECLAGATGS 134
Db 83 DILRCRGKGYEAFLEAFYEPHFLLTUGQEPQRCNSMILDEEGPEGLTFLTEVRR 142
QY 135 LOBELNQERKQEVLLRRCQQLQEHGLGAEATRAEGLHQLEADHSRMRKREVSAPHFVRL 194
Db 143 LREARKSQLQEQQLQARGVLEERAGLEQLRLDQQAQRCQLRREDWEAGSLLELRL 202
QY 195 KDMLSLSLHYNALQEKELASRCRSLQELLYLLKQELQORANNVSSCELELQEQSLRTA 254
Db 203 KENYMIAMRLAOLSEKNSAVLSRSLDQLAVDQLKLVSR--LEEECAL-----LRRA 254
QY 255 SDQESGDEELNRKEENE-----KLRS-----LTFSLAE----- 283
Db 255 RGPPGAEEKEKEKEKEPDNDVLYSELRAENQQLTASLRELQEQLOEASRPGAPGSE 314
QY 284 ---KDILEQSLDARSQELVRIIHSRERAAVAAQREQVWEKEQTLLQFQSKMAC 340
Db 315 RIILLDILEHDWREAQDSRQELCQKLHVAQOGELQWAEELRDQVLOEMEDLRLKHTLQKDC 374

```

```

QY 341 QLYREKVNALQAVCELOKEROQDAYSARDSAQEISQSLEKDSLRROQVVELTDOVCELR 400
Db 375 DLYKHMATVLAQLEIEKERQDAIQSRDRIQIQYSQSLEKDYKQVRGLEADERDELL 434
QY 401 TQLROLQAEPPGVLVQKQEAETR-EPCPREKQRLVRMHATCPRDDSDCSLVST- 451
Db 435 TWTLSLEGTKALLEVLQVLAQGGTCLK-----ACASSHSLCSNLSTWSEPS 484
QY 452 -----ESQLSDLSATSSRELVDSPR-----SSPAPPSSQSLVKRVADGEPR- 496
Db 485 PLGGPEATGEAAVGGGPEPHNSEETADSEKINRLSILPFPFSAGSILRRQEEPPAPK 544
QY 497 WFSFSCLEIPEGDGALPCGAKAGDHPDLYELLDATLPQLESSLOFVSPSRDLVDSESGVL 556
Db 545 RSESSMSDI---TGSV-----TLKPWSPGLSSSSSSDSV 575
QY 557 MRRRPARRILSQVTMLAFOGDALL-----EQISVTGGNLTGFIHR 597
Db 576 WPLGKPEGLLAGCGGLDFLNRSLAIRVSGRSPGPEPDGKPGDGLSFGDRWSGAVRR 635
QY 598 VT--PGSA----ADQMALRPGTQIVMVDYEAEPFLKAVL--EDTTLEBAVGLLRRVDFG 649
Db 636 VLSGPGSARMEPREQVRVEAAGLEGACLEAAQO---RTLWNQGSTLPSL-----MDSK 686
QY 650 CCLSVKVNDDGYKRLLODLEAKVATSG-DSFYIRVNLAMEGRAKGE-LOVHCNEVLHVTD 707
Db 687 ACQSFH-----EALAWAKGPGAEFFYIRANLTLPERADPHALCVKAOEILRLVD 736
QY 708 TFGQCGCWHHRVSVNTKMTAAHGTIPNYSRAOQILALIQDMTQOCTVTRKPS--G 765
Db 737 SAYKRQEWFTCTVDPLTLRD-LDRGTVPNYRAQQL-----EVQEKCL---PSSRHR 786
QY 766 GPQKLVRIVSMD-----KAKASRLRSLDRQGLDPSRMGSSSTCFWAEISCLTLVPTYL 818
Db 787 GPRSNLKKRALDQLRLVRPKVPAGAPAGDSPDQLLEP-----C--APEPSRLRPSYL 836
QY 819 VWHRPAPRPVLLVPRVAVKILSEKLCLLQG----FKKCLAEYLSQEEYAWMSQRGDII 874
Db 837 VRPLLSALRPVLLPCLAPRLIRNLLDLPSSRLDFQVCPAESLSGEECLCPSSAPGAPK 896
QY 875 QEVESSGGRCWVTRHVESLMEKNTHALLDVQDSVCTLHRMDIFPIVIVHVSNEKMAK 934
Db 897 AQPATPGLGSRI-RAIQESVGKK--HCLLELGARGVRVQNEIPIVIVHEVTEKNRE 953
QY 935 LKGLQRLGTSSEOLLEAARQEGDLDRAPCLYSLSLAPDQWSDLOGSLSCVROAIADEQK 994
Db 954 VRGLLGRPGWRDSELLRQCRGSEQVLMLGPCSMVQVPAHEWGHAEELAKVVRGRILQEQ 1013
QY 995 KVVWTE 1000
Db 1014 RLWVVE 1019
QY 1014 RLWVVE 1019

RESULT 6
Q9EPY0 PRELIMINARY; PRT; 536 AA.
ID Q9EPY0
AC Q9EPY0;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CASPASE RECRUITMENT DOMAIN PROTEIN 9.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=20576268; PubMed=11053425;
RA Bertin J., Guo Y., Wang L., Srinivasula S.M., Jacobson M.D.,
RA Poyet J.-L., Merriam S., Du M.Q., Dyer M.J.S., Robison K.E.,
RA DiStefano P.S., Alnemri E.S.;

```

CARD9 is a novel caspase recruitment domain-containing protein that interacts with BCL10/CLAP and activates NF-kappa B.;
J. Biol. Chem. 275:41082-41086(2000).
EMBL: AF311287; AAC28791.1; -.
DR InterPro: IPR001315; CARD.
DR PROSITE: PS50209; CARD; 1.
SQ SEQUENCE 536 AA; 62631 MW; 6F33089CB7E6BAC9 CRC64;

```
Query Match          10.0%; Score 512.5; DB 11; Length 536;
Best Local Similarity 28.8%; Pred. No. 2.2e-22;
Matches 162; Conservative 94; Mismatches 215; Indels 91; Gaps 15;

QY 15 DEETLWMESHRRIRVCIPSRLLTPYLQAKVLCQDEEEVHLSPRLTNSAMRAGHLL 74
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6 NDDCEWSALESFVKLSIVDSPSRITPYLRQCKVLPNDDDEQVLSLPNVIKRVKGVLL 65

QY 75 DLLKTRGKNGAIAFLSKPHNPDPVYTLVTCLOPD-----VDFSNFSGL-----ME 120
   ( : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 DILQRTGHKGVAFALESLEYYPOLYKVTGKEPARVFSMIIDASGESGLTQLLMTVMK 125

QY 121 TSKLTCLAGAGLSQELNOEKGQEVLLRRCQQLQEHGLAETRAEGLHQLLEADHSM 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 126 LQKVVQDLTALLSS-KDDFIKELRVKDSLKKHQRVQ-----RL 164

QY 181 KREVS AHFHEVRLKDEMLSLSHYSNALQEKELAAASCRSLQEBLYLLKQELQRANMVS 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 165 KEECELSSAELKCKDENYDLAMRLAHLSEKGAALMRNLDQLEVDLRHSLMAE--D 222

QY 241 SCELELQEQ-SLRTASQESGDEELNRLKEE-----KLRLSTFSLAEK-----DI 286
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 223 DCKVERKHTLKLHAMEQRPSEQLLDLQOERDLQARVQELVSEVQEGKLRHNSPYIQV 282

QY 287 LEQSLDEARSGRQLVERIHSRLRERAAVAAERQEQYWEKEQTLLQFQKSMACOLYREK 346
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 283 LEEDWRQALQEHQEQASTIFSLRKDLRQAEALRRCWEEKEMFELQCLALRKDKAKMTKDR 342

QY 347 VNALQAVQCELOKQERQAYSARDSAQREISQSLVEKDSLRQVRFELTDQVCELRQLRL 405
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 343 IEAILQMEVSTERQAMTSREELHAQCAQSFQDKDKLRQVRELDEKADELQLQLFQT 402

QY 407 QAEPPGVKQEARTRPCREKORLV---RMHAICPRDDSDCSLVSSTESQSL-LSDLSAT 462
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 403 ESR---LLAAEGRKQ---QQLDMLTLSSDLEDSPPNSQELSLPQLEDAQLSDKGYL 456

QY 463 SSRE-----LVDSFRSSSPAPPSQSLYKRVAEDFG-----EPFWS 498
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 457 ADRESPEQPVVNLKXHLSTHTVTSSEPPPEKERRRLKESFENYRKRALKRMQNSWR 516

QY 499 FSSCLEIPEGDPGALPGAKAGD 520
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 517 QG-----EGDHGNTTGSNDNTD 532
```

```
RESULT 7
QH257 ID Q9H257 PRELIMINARY; PRT; 536 AA.
AC Q9H257
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CASPASE RECRUITMENT DOMAIN PROTEIN 9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20576268; PubMed=11053425;
RA Bertin J., Guo Y., Wang L., Srinivasula S.M., Jacobson M.D.,
RA Poyet J.-L., Merriam S., Du M.Q., Dyer M.J.S., Robison K.E.,
RA Distefano P.S., Alnemri E.S.;
RT *CARD9 is a novel caspase recruitment domain-containing protein that
```

```
RT interacts with BCL10/CLAP and activates NF-kappa B.*;
RL J. Biol. Chem. 275:41082-41086(2000).
DR EMBL: AF311287; AAC28790.1; -.
DR InterPro: IPR001315; CARD.
DR PROSITE: PS50209; CARD; 1.
SQ SEQUENCE 536 AA; 62267 MW; 6EB1835315B83DE5 CRC64;

Query Match          9.6%; Score 493; DB 4; Length 536;
Best Local Similarity 29.5%; Pred. No. 3.2e-21;
Matches 155; Conservative 92; Mismatches 200; Indels 78; Gaps 14;

QY 15 DEETLWMESHRRIRVCIPSRLLTPYLQAKVLCQDEEEVHLSPRLTNSAMRAGHLL 74
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6 NDDCEWNLGEFVRVTLTSVIDPSRITPYLRQCKVLPNDDDEQVLSLPNVIKRVKGVLL 65

QY 75 DLLKTRGKNGAIAFLSKPHNPDPVYTLVTCLOPD-----VDFSNFSGL-----ME 120
   ( : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 DILQRTGHKGVAFALESLEYYPOLYKVTGKEPARVFSMIIDASGESGLTQLLMTVMK 125

QY 121 TSKLTCLAGAGLSQELNOEKGQEVLLRRCQQLQEHGLAETRAEGLHQLLEADHSM 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 126 LQKVVQDLTALLSS-KDDFIKELRVKDSLKKHQRVQ-----RL 164

QY 181 KREVS AHFHEVRLKDEMLSLSHYSNALQEKELAAASCRSLQEBLYLLKQELQRANMVS 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 165 KEECEAGSRELKCKEENYDLAMRLAHLSEKGAALMRNLDQLEIDLQKHSMAE--D 222

QY 241 SCELELQEQ-SLRTASQESGDEELNRLKEE-----NEKRLSTFSLAEK-----DI 286
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 223 DCKVERKHTLKLHAMEQRPSEQLLDLQOERDLQARVQELVSEVQEGKLRDRSPYIQV 282

QY 287 LEQSLDEARSGRQLVERIHSRLRERAAVAAERQEQYWEKEQTLLQFQKSMACOLYREK 346
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 283 LEEDWRQALRHDQEQANTIFSLRKDLRQGEARRLRCEWEEKEMFELQCLALRKDSKMYKDR 342

QY 347 VNALQAVQCELOKQERQAYSARDSAQREISQSLVEKDSLRQVRFELTDQVCELRQLRL 405
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 343 IEAILQMEVSTERQAMTSREELHAQHARGLOEKALRQVRELGEKADELQVQFC 402

QY 407 QAEPPGVKQEARTRPCREKORLV---CPRDDSDCSLVSSTESQSLSD-----458
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 403 EAQ---LLAVEGRUR---QOELTVLSSDLEDSPPRSQELSLPQDLEDQLSDKGLCA 456

QY 459 -----LSATSSRELVSFRS---SSPAPPSSQSLYKRVAEDE 492
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 457 GGSFPKQPFALHQEQVLRNPHDAGLSSGEPPEKER--RRLKESF 499

RESULT 8
QH854 ID Q9H854 PRELIMINARY; PRT; 366 AA.
AC Q9H854
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CDNA FLJ13939 FIS, CLONE Y79AA1000827.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT *NED0 human cDNA sequencing project.*;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK024001; BAB14766.1; -.
```


| | | | |
|----|------|--|------|
| Db | 909 | LARIKLEE-----MDRELPSPSAYDHGD-----NNSISUSETRNV | 947 |
| Qy | 895 | MEKNTHALDVQDSVCTIHRMDIPVIVH-SVNEKMAKKLKGQLRGLTSEBQLEA- | 952 |
| Db | 948 | IAKNKCLLDIVPEGIEMLYAQLCPVYVMLNPSRGAVKDMROSLVK-----EILSSP | 1001 |
| Qy | 953 | -----ARQEEGLDRA-----PCLYSPLAPDCWSDLDGLSCVROAIADEQ | 993 |
| Db | 1002 | TNVTNFTGDPNAKQVKNLFNANKLNEEYPHVFTSKINTHIAGDVL-----SRAFYEKL | 1057 |
| Qy | 994 | KKVVWTEQS | 1002 |
| Db | 1058 | KEVIFTQS | 1066 |

```

RESULT 10
Q9UGR6
ID Q9UGR6 PRELIMINARY; PRT; 332 AA.
AC Q9UGR6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE D3889J22B.1 (NOVEL PROTEIN (ISOFORM 2)) (FRAGMENT).
DE D3889J22B.1.
GN Homo sapiens (Human).
OS
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
RN
RP SEQUENCE FROM N.A.
RA Steward C.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL049451; CAB63075.1; -.
FT non_ter
SQ SEQUENCE 332 AA; 37053 MW; 68234A527D009007 CRC64;

```

Query Match 6.1%; Score 315; DB 4; Length 332;
Best Local Similarity 31.4%; Pred. NO. 5.3e-11;
Matches 106; Conservative 51; Mismatches 145; Indels 36; Gaps 10;

| | | |
|----|-----|--|
| Qy | 677 | DSFVIRVNLAMEGRAGE-IQVHCNEVLHVDTDFMGGCGWHAHRVNSVTMKDTAAHGPI 7335 |
| Db | 4 | EPFYIRANLTPERADPHALCVKAQKELRLVDSAYKRRROEWFCTRPVDPLTLRLD-LDRGTV 62 |
| Qy | 736 | PNTYSRAQQOIALIQQMTQOCTVTRPSS--GGPQKLVLRIVSMD-----KAKASPLRL 786 |
| Db | 63 | PNTYQRAOQLL-----EYQEKCL----PSSHRGPRLNKKRALDQLURLVRRPVPVGPAGAD 113 |
| Qy | 787 | SFDRGQLDPSRMESGSCFWAESCLTLVPYTLVMPHRPARPRPVLLVPRVAVGKILSEKLC 846 |
| Db | 114 | SPDQLLLEP-----C--AEPERSUYSVLVRPLLVLSALRPVVLVLPCLAPRLIRNL 163 |
| Qy | 847 | LLQG-----FRKCLAEYLSQBEYEAWSQRGDIIOEGEVSGRGCVWTRHVAVESLMEKNTHAL 902 |
| Db | 164 | DLPSRLDFQVCPAESLSGELCPSSAPGAPKAQPATPGLGSRI--RAIQESVGRQKHCL 221 |
| Qy | 903 | LDVOLDSCVTLHRMDTFPIVHVSVNEKMAKKLKGQLRGLTSEBOLLEAARQEEGDLDR 962 |
| Db | 222 | LELGARGVRELQNEIYPIVITHVEVTEKNVREVRGLLGRPGWRNDSSELLRQCRGSEQVLWG 281 |
| Qy | 963 | APCLYSSLADPGHSDDLLGLLSCVRQAIADQKKVWTE 1000 |
| Db | 282 | LPCSWGVQPAHEHGHAEEELAKVYVRGRIQEARLVWVE 319 |

| | |
|---------|---|
| RESULT | 11 |
| Q9UGR5 | |
| ID | Q9UGR5 |
| AC | PRELIMINARY; PRT; 331 AA. |
| Q9UGR5; | |
| DT | 01-MAY-2000 (TReMBLrel. 13, Created) |
| DT | 01-MAY-2000 (TReMBLrel. 13, Last sequence update) |

| | |
|----|--|
| DT | 01-MAY-2000 (T-EMBLrel. 13, Last annotation update) |
| DE | DJ889022B.1 (NOVEL PROTEIN (ISOFORM 1)) (FRAGMENT). |
| DN | DJ889022B.1. |
| GC | Homo sapiens (Human). |
| OS | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| OX | NCBI_TaxID=9606; |
| EN | [1] |
| RP | SEQUENCE FROM N.A. |
| RA | Steward C.; |
| RL | Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases. |
| RD | EMBL; AL049851; CAB63076.1; -. |
| FT | NON_TER 1 |
| SO | SEQUENCE 331 AA; 36925 MW; DFF0A6323E39F65E CRC64; |

Query Match 6.0%; Score 307.5; DB 4; Length 331;
Best Local Similarity 32.0%; Pred. No. 1.5e-10;
Matches 108; Conservative 50; Mismatches 143; Indels 37; Gaps 11;

| | | | |
|----|-----|--|------|
| Qy | 677 | DSFYIRVNLAMEGRKAGE--LQVHCNEVHLVTDTFMGCQCWHAHRYVNSYTMKDTAAHGFI | 7335 |
| Db | 4 | EPFYIRANLTLPERADPHALCVKRAQELRLVDSAYKKRQEWECTFRVDPDLTLD--LDGRGV | 62 |
| Qy | 736 | PNYSRAQOOLIALIQDWTQCTVTRPSS--GGPKQLRVISMD-----KAKASPLRL | 786 |
| Db | 63 | PNYORAQOLL-----EYQEKL----PSSRHRRPSRLNKKRALDQLRLVRPVPVGPAGD | 113 |
| Qy | 787 | SFDRGOLDPSRMGSSTCFNAESCLTLPVTLVNPHRPARPRPVLLVPRAVGKILSEKLC | 846 |
| Db | 114 | SPDQLLLEP-----C--AEPERSLPYSVRLVLSALRPVLLPECLAPRLTRNL | 163 |
| Qy | 847 | LLQG-----FKKLAELYSQEYEAWSQGDIIIEGEVSGRCWTVRHAYESLMKNTHAL | 902 |
| Db | 164 | LPSSRLDFQCPAESLSGELCPSSAPGAPKAPQATPGLGSR--RAIQESVGKK--HCL | 220 |
| Qy | 903 | LDVQLDSVCTLHRRMDIEPIVIHVSNEKMAKLLKGLQLRGLTSEEQLLEAAARQEEGDLR | 962 |
| Db | 221 | LELGARGVRELQVNEIPIVIHVTEKTVREYVRLGLLRPGWRDSELLRQCRGSEQVLWG | 280 |
| Qy | 963 | APCLYSLPADGWSDDLGLLSCVRQAIADQQKVVWTE | 1000 |
| Db | 281 | LPCSWGVQPAHEWGHAELAKVYVGRILQIQEARLVWVE | 318 |

| | | |
|--------|-------------|---|
| RESULT | 12 | |
| O97758 | | |
| ID | O97758 | PRELIMINARY; PRT: 1769 AA. |
| AC | O97758; | |
| DT | 01-MAY-1999 | (TREMBLrel. 10, Created) |
| DT | 01-MAY-1999 | (TREMBLrel. 10, Last sequence update) |
| DT | 01-DEC-2001 | (TREMBLrel. 19, Last annotation update) |
| DE | ZO-1 MDCK | |

OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
ON NCBI_TaxID=9615;
RX [1]
RN SEQUENCE FROM N.A.
RP STRAIN-BREED: COCKER SPANIEL;
RC MEDLINE=99196918; PubMed=10094817;
RX Gonzalez-Mariscal L., Islas S., Contreras R.G., Garcia-Villegas M.R.,
RA Betanzos A., Vega J., Diaz-Quinonez A., Martin-Orozco N.,
RA Ortiz-Navarrete V., Cerejildo M., Valdes J.;
RT "Molecular characterization of the tight junction ZO-1 in MDCK
cells.";
RL Exp. Cell Res. 248:97-109(1999).
DR ENBL; U55935; AAD11529.1; -.
DR HSP; P31016; IBE9.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR

Fri Jun 14 08:13:31 2002

| | |
|----|--|
| DR | InterPro: IPR000906; ZU5. |
| DR | Pfam: PF00525; Guanylate_kin; 1. |
| DR | Pfam: PF00695; PDZ; 3. |
| DR | Pfam: PF00018; SH3; 1. |
| DR | Pfam: PF00791; ZU5; 1. |
| DR | SMART: SM00072; GUKc; 1. |
| DR | SMART: SM00228; PDZ; 3. |
| DR | SMART: SM00326; SH3; 1. |
| DR | SMART: SM00218; ZU5; 1. |
| DR | PROSITE: PS50052; GUANYLATE_KINASE_2; 1. |
| DR | PROSITE: PS50106; PDZ; 3. |
| DR | PROSITE: PS50002; SH3; 1. |
| SQ | SEQUENCE 1769 AA; 197606 MW; 181E9F36CEBC96EF CRC64; |

| | | |
|---------------------------|--|---|
| Query Match | | 5.6%; Score 290.5; DB 6; Length 1769; |
| Best Local Similarity | | 23.6%; Pred. No. 1.4e-08; |
| Matches 188; Conservative | | 121; Mismatches 288; Indels 201; Gaps 41; |

| | | |
|----|-----|---|
| Qy | 247 | QEQLRTASDQESGDEE-----LNRLKEENEKLRSTFTSLAEKDILEQSDLEARG 296 |
| Dd | 162 | KERSLSPPDRRSVASSOPPKTKVTLVKSRKNEYGLR-LASHIFVEISODSLSAARDG 220 |
| Qy | 297 | SRQE--LVERIH-SLRERAVAAERQYEEKEOTLLQFOKSKMACOLYRKVNALQAQ 353 |
| Dd | 221 | NIQEGDVVLKINGTVTNNSLTD-----AKTLIERSKGKLKVVQRDE----RAT 266 |
| Qy | 354 | VELOKERDAQYSARDSAQRETS--QSLVKDSLRRQVFELTDVCVELRTLQRLQAEP 411 |
| Dd | 267 | LUNVPDLSDSHSANASERDDISEIQSLASDHGRS-----HDRPP 307 |
| Qy | 412 | GVLKQEAHTREPCKEORVLRMHAICPRDDSDCLSVLSTESQLLSDLSATSRELVD 471 |
| Dd | 308 | ----RHRSRSPDQRSEPSDHSRHS-PQQPSSGSL-RSREERIISKPGAV----- 351 |
| Qy | 472 | RSSPAPPSQQSLYKRVAEDFGEEFPWFSFSCLEIEPEGDPGALPGAAGPHLDYELLDTA 531 |
| Dd | 352 | --STPVKHADHDTTHVEEVVYER--NEKAQPSLPPEKP--VYAQVGQDFVDL----- 398 |
| Qy | 532 | DLPOLESSLQPVSPP--GRUDVS--BSGVLMRRRPARRILSQVTMLAFQGDALLEQISVIG 587 |
| Dd | 399 | -----PVSPDCGLPNSTHEDGIL--RPSMKLYK----FRKGDSV--GLRLAG 438 |
| Qy | 588 | GNLTGIFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFAVLEDITL-----EEA 639 |
| Dd | 439 | GNDVGIFVAGVLSDSPAAREGLEEGDIILRVN---NVDEFNTIREEAVLFLLDLPRKEEV 495 |
| Qy | 640 | VGLLRRVDGFCCLSVKVNTDGKYLRLQDLLEAKVATSGDSFYIRVNLAAMEGRAKGE 699 |
| Dd | 496 | TILAQK-----KKDVRIIVE-----SDVGDSFYIRTHEYEKESPYGLSFNK 538 |
| Qy | 700 | NEYLHVTDYTMFQG-CGCCHAHRVNSYTMKDTAAHGTPINYSRAQQOLIALIQTMOCTV 758 |
| Dd | 539 | GEVFRVVDLTLYNGKLSWLAIIRGKNHKE--VERGIIPKNRAEQ--LASVQ-----Y 587 |
| Qy | 759 | TRKPSGGQPKLVIRIVSMDKAKASPLR-LSFDRGOLDPSRWEGSSTCFWAESCCLTLPV-Y 816 |
| Dd | 588 | TLPKTAGG-----DRADFWRFERGLRSSRNURKKSREDLS-----AQPVQTFPAY 632 |
| Qy | 817 | TLVMPHRPARPRPVLLVPRAVGKILSEKCLLQGFKKLAEYLSOEI---YE-AWSQRGD 872 |
| Dd | 633 | ERVVLREAGFLRPVTF-GPIADVAREK-----LAREPDYVIQAKSEPRD 677 |
| Qy | 873 | IIOGEVSGRCVWVTRHAVSELMKENTHALLDVQSDVCTLHRMDIPPIVLHVSVMKMA 932 |
| Dd | 678 | AGTDQRSRG---IIRLTIKIQIIDKHALDVTNPNAVRLNLTVAQWYPIVVFVLPDSPKQG 734 |
| Qy | 933 | KKL-----KKGLORLGTSBOLLLEAARQEBGDLDRAPCIYSSLPADGWSDLDGLS 983 |
| Dd | 735 | VKTMRMRCLCPESRKSAKLVERS HKL-----RKNHHLLFTTINLSNM-NDGW-----YG 783 |
| Qy | 984 | CVROIADAEOKKVVYWTQ 1001 |

| | |
|---|---|
| DR | EMBL; D83477; BA11923.1; -. |
| DR | HSSP; Q12923; 3PDZ. |
| DR | FlyBase; FBgn0003177; pyd. |
| DR | InterPro; IPR000619; Guanylate_kin. |
| DR | InterPro; IPR001478; PDZ. |
| DR | InterPro; IPR001452; SH3. |
| DR | Pfam; PF00625; Guanylate_kin; 1. |
| DR | Pfam; PF00595; PDZ; 3. |
| DR | SMART; SM00072; GUKC; 1. |
| DR | SMART; SM00228; PDZ; 3. |
| DR | SMART; SM00326; SH3; 1. |
| DR | PROSITE; PS50052; GUANYLATE_KINASE_2; 1. |
| DR | PROSITE; PS50106; PDZ; 3. |
| SQ | SEQUENCE 1367 AA; 148347 MW; 571C4566C6B68BF8 CRC64; |
| Query Match 5.5%; Score 282; DB 5; Length 1367; | |
| Best Local Similarity 22.7%; Pred. No. 3.1e-08; | |
| Matches 155; Conservative 111; Mismatches 276; Indels 142; Gaps 30; | |
| Qy | 373 REISQSLVEKDSLRQVFELTDQVCELRQLQLOAEPFGVLKQEAETREPCPREKORLV 432 |
| Db | 194 KEISSK--AREQLNANGYSL--QEGDIITRIHNTNCGDTMSLKEAKKIIDGC--KERLNLV 248 |
| Qy | 433 RMHAICPRDD--SDCSLVSTESQLLSLSAT-----SSRELVDSPRSSSPAPPQ 481 |
| Db | 249 VLRLITQAVSQNLNANNSASHQASGNIYATHQPOVSGCSCSSNNNLEDPYLPGG--ASYSS 307 |
| Qy | 482 QSLYKRVAEDEGEWPWFSSCLETPEGDGALPGAKAGDHPHLDYELLDYADLPQLESSLQ 541 |
| Db | 308 QNLVQPPTRTSNGPNTNGNLN--DEKSNLTPRGRSRGP-----IMDGVSLQQLD---R 357 |
| Qy | 542 PVSPCR-----LDV-----SESGVLM-----RRRPARILSOVTMLAFQGDALL 580 |
| Db | 358 PVTPTGRSAADIEPPRPDPGSGGAAQDFYSSRQLYEERQSAEPRFISFKEGSV 417 |
| Qy | 581 EQLSVIGNLGTFIHRVTGCSAADQMALRPGTQIVMVYDEASEPLFKAVLEDTTLEEAV 640 |
| Db | 418 -GIRLTGGNEAGIFVAVQPCSPASLQGLMPGDKILVN-----DMDMNGVTREEAV 468 |
| Qy | 641 GLLRVDFGCLSVKVNNTDGYKRLLODLEAKVAT--GDSFYIRVNLAMEGRKAGELQVH 698 |
| Db | 469 LFL-----LSLQDRIDLIVQYCKEYDEVVTNQRGDSPHIRTHCDNPSKGEAFK 520 |
| Qy | 699 CNEVLHVTDTMFQ--CGCWAHRYNSYTMKDTAAHGTPINYSRAQQQLIALIODMTQOCT 757 |
| Db | 521 AGDVFVVIDTLHNGVVGWQVLKIGRHQE--MQGVIPNKSRAEEL-----ATAQFN 571 |
| Qy | 758 VTRKPSGGPQKLVRIVSMOKAKASPLRLSFDRGOLDPSR-----MEGSSTCFWAESCL 811 |
| Db | 572 ATKK-----EMNANESRGNFRRRRSTRHSKSLSRENWDVDFSDSIS 615 |
| Qy | 812 TLVPVTLVWPHRPARPRPVLLVPRAVGKILSEKLCQLQGFKKCLAEYLSQEEYEAWSQRG 871 |
| Db | 616 KPPATERVVLRHPGFRVPVLF--GPVSLARERLA--KDFDPKFSTPLQDDDKSA-----667 |
| Qy | 872 DTIQEGEVSGGRCWVTRHA--VESLMEKNTHALLDVQLDSVCTLHRMDIFPIVHVSVNEK 930 |
| Db | 668 -----ATSGKCRIVRLSNIRDVMDRGKHALLDITPNAVDRLNYAQFYPVVFILKTDK 720 |
| Qy | 931 -MAKKLKGLQRLG--TSEQLLEAARQEGDLDRAPCLYSSLPDGV-----SDL 978 |
| Db | 721 HVIKQLRHGLPKAAHKSSKKLLEOCQK-----LERV-----WSHFSTQIALSDE 765 |
| Qy | 979 DGLLSCVROAIADEQKVVWTEQS 1002 |
| Db | 766 ESWYKLRDSIDLQSGAVWMSSES 789 |

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2002, 09:22:57 ; Search time 15.16 Seconds
(without alignments)
2564.276 Million cell updates/sec

Title: US-09-767-215-2

Perfect score: 5149

Sequence: 1 MGELCRRDSALTALDEETLW.....VROAIADQKVVWTEQSPR 1004

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query % | Length | ID | Description |
|------------|--------|---------|--------|--------------|---------------------|
| 1 | 5149 | 100.0 | 1004 | 1 CARE_HUMAN | Q9bx16 homo sapien |
| 2 | 3869.5 | 75.2 | 999 | 1 CARE_MOUSE | Q99kfo mus musculus |
| 3 | 1239.5 | 24.1 | 1147 | 1 CARB_HUMAN | Q9bx17 homo sapien |
| 4 | 887.5 | 17.2 | 1032 | 1 CARB_HUMAN | Q9bwt7 homo sapien |
| 5 | 886.5 | 17.2 | 1021 | 1 CARA_MOUSE | P58660 mus musculus |
| 6 | 512.5 | 10.0 | 536 | 1 CAR9_RAT | Q9epv0 rattus norv |
| 7 | 493 | 9.6 | 536 | 1 CAR9_HUMAN | Q9b257 homo sapien |
| 8 | 289 | 5.6 | 1745 | 1 ZO1_MOUSE | P39447 mus musculus |
| 9 | 280 | 5.4 | 1167 | 1 ZO2_MOUSE | Q920u1 mus musculus |
| 10 | 270.5 | 5.3 | 1736 | 1 ZO1_HUMAN | Q07157 homo sapien |
| 11 | 256 | 5.0 | 1174 | 1 ZO2_CANFA | Q95168 canis faml |
| 12 | 254.5 | 4.9 | 1190 | 1 ZO2_HUMAN | Q9udy2 homo sapien |
| 13 | 249 | 4.8 | 4684 | 1 PLE1_HUMAN | Q15149 homo sapien |
| 14 | 246.5 | 4.8 | 905 | 1 ZO3_MOUSE | Q9qxy1 mus musculus |
| 15 | 240.5 | 4.7 | 1960 | 1 MYH9_HUMAN | P35579 homo sapien |
| 16 | 238.5 | 4.6 | 898 | 1 ZO3_CANFA | O62683 canis faml |
| 17 | 236.5 | 4.6 | 4473 | 1 PLE1_CRIGR | Q9ji55 cricetus |
| 18 | 236 | 4.6 | 3210 | 1 CENF_HUMAN | P49454 homo sapien |
| 19 | 235 | 4.6 | 1959 | 1 MYH9_CHICK | P14105 gallus gall |
| 20 | 234.5 | 4.6 | 4687 | 1 PLE1_RAT | P30427 rattus norv |
| 21 | 226.5 | 4.4 | 1976 | 1 MYHA_RAT | Q9j1t0 rattus norv |
| 22 | 226 | 4.4 | 1972 | 1 MYHB_RABIT | P35748 oryctolagus |
| 23 | 225.5 | 4.4 | 2349 | 1 TPR_HUMAN | P12270 homo sapien |
| 24 | 225.5 | 4.4 | 2704 | 1 BPAL_HUMAN | Q03001 homo sapien |
| 25 | 223 | 4.3 | 1939 | 1 MYH1_HUMAN | P12882 homo sapien |
| 26 | 219 | 4.3 | 1972 | 1 MYHB_HUMAN | P35749 homo sapien |
| 27 | 218 | 4.2 | 2230 | 1 GOG4_HUMAN | Q13439 homo sapien |
| 28 | 217.5 | 4.2 | 933 | 1 ZO3_HUMAN | O95049 homo sapien |
| 29 | 217.5 | 4.2 | 1976 | 1 MYHA_HUMAN | P35580 homo sapien |
| 30 | 217 | 4.2 | 1972 | 1 MYHB_MOUSE | O08638 mus musculus |
| 31 | 215.5 | 4.2 | 1087 | 1 AKA9_RABIT | P28628 oryctolagus |
| 32 | 214 | 4.2 | 1407 | 1 TRHY_RABIT | P37709 oryctolagus |
| 33 | 214 | 4.2 | 1941 | 1 MYH2_HUMAN | Q9ukx2 homo sapien |

| | | | | | |
|----|-------|-----|------|--------------|---------------------|
| 34 | 212.5 | 4.1 | 1940 | 1 MYH3_RAT | P12847 rattus norv |
| 35 | 212 | 4.1 | 1957 | 1 YD86_SCHPO | Q10411 schizosacch |
| 36 | 212 | 4.1 | 1976 | 1 MYHA_BOVIN | Q27991 bos taurus |
| 37 | 211.5 | 4.1 | 1939 | 1 MYH4_HUMAN | Q9y623 homo sapien |
| 38 | 210.5 | 4.1 | 1976 | 1 SCPI_HUMAN | Q15431 homo sapien |
| 39 | 209.5 | 4.1 | 1937 | 1 MYH8_HUMAN | P13535 homo sapien |
| 40 | 209 | 4.1 | 1938 | 1 MYHD_HUMAN | Q9ukx3 homo sapien |
| 41 | 207.5 | 4.0 | 1790 | 1 USO1_YEAST | P25386 saccharomyc |
| 42 | 207.5 | 4.0 | 2871 | 1 DESP_HUMAN | P15924 homo sapien |
| 43 | 207 | 4.0 | 1084 | 1 MYSS_RABIT | P02562 oryctolagus |
| 44 | 207 | 4.0 | 1325 | 1 GI60_MOUSE | P55937 mus musculus |
| 45 | 206.5 | 4.0 | 1290 | 1 XCPC_XENLA | P50532 xenopus lae |

ALIGNMENTS

```
RESULT 1
CARE_HUMAN
ID CARE_HUMAN STANDARD; PRT; 1004 AA.
AC Q9BX16: Q9BV55;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE DE Caspase recruitment domain protein 14 (CARD-containing MAGUK protein
DE 2) [Carma 2].
GN CARD14 OR CARMA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21192234; PubMed=11278692;
RA Bertin J., Wang L., Guo Y., Jacobson M.D., Poyet J.-L.,
RA Srinivasula S.M., Merriam S., Distefano P.S., Alnemri E.S.;
RT "CARD11 and CARD14 are novel caspase recruitment domain
RT (CARD)/membrane-associated guanylate kinase (MAGUK) family members
RT that interact with Bcl10 and activate NF-kappaB.";
RL J. Biol. Chem. 276:11877-11882(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX PubMed=11356195;
RA Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;
RT "Carmal, a CARD-containing binding partner of Bcl10, induces Bcl10
RT phosphorylation and NF-kappaB activation.";
RL FEBS Lett. 496:121-127(2001).
RN [3]
RP ERRATUM.
RA Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;
RL FEBS Lett. 505:198-198(2001).
RN [4]
RP SEQUENCE OF 1-740 FROM N.A.
RC TISSUE=Cervix, and Colon;
RA Strausberg R.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK. Stimulates the
CC phosphorylation of Bcl10.
CC -!- SUBUNIT: CARD14 and Bcl10 bind to each other by CARD-CARD
CC interaction.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: Expressed in placenta. Also detected in HeLa
CC S3 cells, but not in the other cancer cell lines tested.
CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- CAUTION: Supposed to contain a SH3 domain which is not detected by
CC PROSITE, Pfam or SMART.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

| | | | | | | | | | | | | | | |
|-----|---------|--------|--------|-------|--------|-------|-------|--------|--------|------|------|-------|-----|-----|
| 1 | MGELCRD | SALTAL | DEETLW | EMESH | RHVRIC | PSRPT | YLRLQ | AKVLCQ | DEBEVL | HS | 60 | | | |
| 1 | MGELCRD | SALTAL | DEETLW | EMESH | RHVRIC | PSRPT | YLRLQ | AKVLCQ | DEBEVL | HS | 60 | | | |
| 61 | PRLTNSA | RAGHL | DLKTR | GKN | GAFA | LES | KFN | PDVY | TLTGL | OPDV | SFNS | SGLME | 120 | |
| 61 | PRLTNSA | RAGHL | DLKTR | GKN | GAFA | LES | KFN | PDVY | TLTGL | OPDV | SFNS | SGLME | 120 | |
| 121 | TSKLT | ECLAG | ISLQ | BE | LNQ | EGK | QEV | LLRR | CQ | Q | Q | Q | Q | 180 |
| 121 | TSKLT | ECLAG | ISLQ | BE | LNQ | EGK | QEV | LLRR | CQ | Q | Q | Q | Q | 180 |
| 181 | KREYSA | HFV | RLK | DE | ML | SL | SHY | SNA | LQ | EK | E | L | A | 240 |
| 181 | KREYSA | HFV | RLK | DE | ML | SL | SHY | SNA | LQ | EK | E | L | A | 240 |
| 241 | SCELE | Q | ES | L | UR | TAS | D | ES | G | DE | E | L | N | 300 |
| 241 | SCELE | Q | ES | L | UR | TAS | D | ES | G | DE | E | L | N | 300 |
| 301 | LVERI | H | S | L | R | E | R | A | V | A | E | R | Q | 360 |
| 301 | LVERI | H | S | L | R | E | R | A | V | A | E | R | Q | 360 |
| 361 | RDQAY | S | A | R | D | S | A | Q | R | E | I | S | O | 420 |
| 361 | RDQAY | S | A | R | D | S | A | Q | R | E | I | S | O | 420 |
| 421 | REPCP | R | E | Q | R | L | V | R | H | A | I | C | P | 480 |
| 421 | REPCP | R | E | Q | R | L | V | R | H | A | I | C | P | 480 |
| 481 | QOSLY | K | R | V | A | E | D | F | G | E | E | P | N | 540 |
| 481 | QOSLY | K | R | V | A | E | D | F | G | E | E | P | N | 540 |
| 541 | OPVSP | G | R | L | D | V | S | E | G | V | L | M | R | 600 |
| 541 | QVPSG | R | L | D | V | S | E | G | V | L | M | R | R | 600 |

```

01 - MAR-2002 (Rel. 41, Last sequence update)
01 - MAR-2002 (Rel. 41, Last annotation update)
DE DE caspase recruitment domain protein 14 (Bcl10-interacting MAGUK protei
DE DE 2) (Bimp2).
GN CARD14 OR BIMP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21391892; PubMed=11387339;
RA McAllister-Lucas L.M., Inohara N., Lucas P.C., Ruland J., Benito A.,
RA Li Q., Chen S., Chen F.F., Yamaoka S., Verma I.M., Mak T.W.,
RA Nunez G.;
RT "Biml1, a MAGUK family member linking protein kinase C activation to
RT Bcl10-mediated NF-kappa B induction.";
RL J. Biol. Chem. 276:30589-30597(2001).
RN [2]
RP SEQUENCE OF 82-743 FROM N.A.
RC TISSUE=Breast;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/Genbank/DDBJ databases.
CC -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK.
CC -!- SUBUNIT: CARD14 and Bcl10 bind to each other by CARD-CARD
CC interaction (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DRH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- CAUTION: Supposed to contain a SH3 domain which is not detected
CC PROSITE, Pfam or SMART.

```

CC This SWISS-PROT entry is copyright, it is protected under copyright law and is the property of the EMBL outsta
CC between the Swiss Institute of Bioinformatics and the EMBL outsta
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in n
CC modified and this statement is not removed. Usage by and for comm
CC entities requires a license agreement (see <http://www.isb-sib.ch/ann>

or send an email to license@isb-sib.ch).

```

CC -----
CC EMBL; AF363457; AAK60137.1; -.
CC EMBL; BC004692; AAH04692.1; -.
CC InterPro; IPR001478; PDZ.
CC Pfam; PF00595; PDZ; 1.
CC SMART; SM00228; PDZ; 1.
CC PROSITE; PS0209; CARD; 1.
CC PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
CC PROSITE; PS00552; GUANYLATE_KINASE_2; 1.
CC PROSITE; PS0106; PDZ; 1.
CC Colled coil. 15 107 CARD.
CC DOMAIN 125 411 COILED COIL (POTENTIAL).
CC FT DOMAIN 125 411 PDZ.
CC FT DOMAIN 572 655 PDZ.
CC FT DOMAIN 854 986 GUANYLATE_KINASE.
CC FT CONFLICT 736 743 QAOQLLA -> HLLEDHRS (IN REF. 2).
CC SEQUENCE 999 AA; 113496 MW; D18350DA12430255 CRC64;

Query Match 75.2%; Score 3869.5; DB 1; Length 999;
Best Local Similarity 76.9%; Pred. No. 4e-186;
Matches 774; Conservative 83; Mismatches 138; Indels 11; Gaps 7;

Qy 1 MGELCRDSALTALDEETLWEMESHRIIVRCICPSRLTPYLROAKVLCOLDEEVLHS 60
Db 1 MAELCRMDSTLTALDEEMLDWLESHRCHRIQVSCPSRLTPYLROAKVLCOLDEEVLHS 60
Qy 61 PRLNLSAMRAGHLDDLKTRCKNGAIAFLSKPHNPDPVYTLVTCLOPDVDFSNFSGLME 120
Db 61 SRFTNSAMRVGHLDDLKARCKNGAIAFLSKPHNPDPVYTLVTCLOPDVDFSNFSGLME 120
Qy 121 TSKITECLAGAGISQLEELNKGQKVEYLLRRCCQLQELHGLAETRAEGLHQLADHGRM 180
Db 121 TSKITECLAGAGISSQLEELAKQKVEYLLRRCCQLKELGLAEAHESGLQLEVDHGRM 180
Qy 101 KREVSAPHEVRLKDEMLSLSHYSNALQEKELAAASCRSLQELLYLLKQLQRANVVS 240
Db 101 KREVSAPHEVRLKDEMLSLSHYSNALREKELAAATRSCHSLQELLYLVKQLQRASIVS 240
Qy 241 SCELELOQSRTASQDSGDEELNRLKEENKRLSLTFLSAENKDLFQSLDEARGSQE 300
Db 241 SCERESRSLKMASNLEPQSGDEELNRLKEENKRLSLTFLSAENKDLFQSLDEARGSQE 300
Qy 301 LVERIHLRERAVAAERQYWEKEQTLQFQKMACOLYREKVNALQAQVCELQKE 360
Db 301 LVDRHLSLRERAVAAERQYWEKEQTLQFQKMACOLYREKVNALQAQVCELQKE 360
Qy 361 RDQAYSARDSAQREISQSLVKDSLRRQVFEITDQVCELTQLRQLQAEPPGVKQEAR 420
Db 361 RDQAYTARDQAQMETSORLVEKDALRRRVFEITDQVCELTQLRQLQAEPPGVKQEAR 420
Qy 421 REPCREKQRLVRHAIACPRDSDCSLVSTFESQLLSATSSRELVDSPSSPAPPS 480
Db 421 RELCLURQRLVRHAIACPRDSDCSLVSTFESRLWDLNSTSSREQMDSFRSSPMPPS 480
Qy 481 QQSILKRYAEDFGEPPWFSFSCLEIPEGDGALPGAKAG--DPLHLYELLDTADLPQLES 538
Db 481 QQSILKRYAEDFLEDPES----LSFPEVLEMLRQATVDDTDTLEFEMIDGADLSQTED 536
Qy 539 SLQVSPGRDLVSSGVLMRRRARRILISQVMTLAFQGDALLEISVIGGNLTGFIHVR 598
Db 537 SLQG--SSRSLNVSSESVVRRRARRKILISQVMTLAFQGDALLEISVIGGNLTGFIHVR 595
Qy 599 TPGSAAQDMLRPGTQIVMDVYESEPLFKAVLEDTTLEAAVGLRRYDVGCCSLSVKNT 658
Db 596 TPGSAAQDMLRPGTQIVMDVYESEPLFKAVLEDTTLEAAVGLRRYDVGCCSLSVKNT 655
Qy 659 DGYKRLQLDLAKVATGSDSFYIRVNLAMEGRAGKELQVHCNEVLHVTDTMFQGCSCWHA 718
Db 656 EGYKNLIQDLAKVATGSDSFYIRVNLAMEGRAGKELQVHCNEVLHVTDTMFQGCSCWHA 715
Qy 719 HRVNSYTKMKTAAHGFTIPNYSRAOQQLLALIQDMTQOCTVTRKPSGGPGKLVIRIVSMDK 778

```

```

Db 716 HNVNPTTKMEP-CTIPNYSQAQQLLALIQDMTQOCTVTRKPK-PGGPKLVIRIVSVDK 773
Qy 779 AKASPLRLSFDGQOLDPSRMEGS-STCFWAESCLILVPYTLVWPHRPARPRVLLVPRAY 837
Db 774 AAVSPLTSFDFQSQDWSGKEGGVSCFWSSECTFLAPYTLVHPRPARPRVLFVPRLV 833
Qy 838 GKILSEKLCILQGFKKCLAEYLSQEEYEAWSQSGDIIQEGEYSGGRCWYTRHVESLMEK 897
Db 834 GRILGKLCILQGFKKCLAEYLSQEEYEAWSQSGDIIQEGEYSGGRCWYTRHVESLMMN 893
Qy 898 NTHALLDVLDSVCTLHRMDIPPIVHVSVNEMAKKLGKLGRLGTSEQLLEAARQEE 957
Db 894 STHALLDVLDSVCTLHRMDIPPIVHVSVNEMAKKLGKLGRLGTSEQLLEAARQEE 953
Qy 958 GDLDRAPCLYSLAPDGSGLDGLLSCVROATADEQKVVWTEQSP 1003
Db 954 GELDRVPCLYSLAPDGSGLDGLLSCVRLAIADEQKVVWTE-SP 998

RESULT 3
CARB_HUMAN
ID CARB_HUMAN STANDARD; PRT: 1147 AA.
AC Q9BXL7;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Caspase recruitment domain protein 11 (CARD-containing MAGUK protein
DE 3) (Carma 1)
DE CARD11 OR CARMA1.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21192234; PubMed=11278692;
RA Bertin J., Wang L., Guo Y., Jacobson M.D., Poyet J.-L.,
RA Srinivasula S.M., Merriam S., DiStefano P.S., Alnemri E.S.;
RT "CARD11 and CARD14 are novel caspase recruitment domain
RT (CARD)/membrane-associated guanylate kinase (MAGUK) family members
RT that interact with Bcl10 and activate NF-kappaB.";
RL J. Biol. Chem. 276:11877-11882(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX PubMed=11356195;
RA Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;
RT "Carmal, a CARD-containing binding partner of Bcl10, induces Bcl10
RT phosphorylation and NF-kappaB activation.";
RL FEBS Lett. 496:121-127(2001).
RN [3]
RP ERRATUM.
RA Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;
RL FEBS Lett. 505:198-198(2001).
CC -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK. Stimulates the
CC phosphorylation of Bcl10.
CC -!- SUBUNIT: CARD11 and Bcl10 bind to each other by CARD-CARD
CC interaction.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: Detected in adult peripheral blood leukocytes,
CC thymus, spleen and liver. Also found in promyelocytic leukemia HL-
CC 60 cells, chronic myelogenous leukemia K562 cells, Burkitt's
CC lymphoma Raji cells and colorectal adenocarcinoma SW480 cells. Not
CC detected in HeLa S3, Molt-4, A549 and G431 cells.
CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -!- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC -!- CAUTION: SUPPOSED to contain a SH3 domain which is not detected by
CC PROSITE, pfam or SMART.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: AF322641; AAG53402.1; -
InterPro: IP001478; PDZ.
Pfam: PF00595; PDZ; 1.
SMART: SM00228; PDZ; 1.
DR PROSITE; PS0209; CARD; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR PROSITE; PS00502; GUANYLATE_KINASE_2; FALSE_NEG.
DR PROSITE; PS50106; PDZ; FALSE_NEG.
KW Coiled coil. 11 103 CARD.
FT DOMAIN 123 442 COILED COIL (POTENTIAL).
FT DOMAIN 123 442 PDZ.
FT DOMAIN 673 748 GUANYLATE KINASE.
FT DOMAIN 966 1133 GUANYLATE KINASE.
FT CONFLICT 808 808 P -> L (IN REF. 2).
SQ SEQUENCE 1147 AA; 132641 MW; 913A4B015D2B36CC CRC64;

Query Match 24.1%; Score 1239.5; DB 1; Length 1147;
Best Local Similarity 30.5%; Pred. No. 9.5e-55;
Matches 354; Conservative 204; Mismatches 402; Indels 201; Gaps 33;

Qy 15 DEETLWEMSHRRHVRICPCSLTPYLQAKVLCOLDEEVLHSPRLNSAMRAGHLL 74
Db 11 EEDALWENVECNHRLSRYINPAKLTPLQCKVIDQDEVLNAPMLSKINRAGRLL 70
Qy 75 DLLTRKNGAIAFLSKLHFNPDVYTLVGLQPDVDFSNFSGLMETSFKTECLAGAGS 134
Db 71 DILHTKGQGVVFLSLEFYPELYKLVTGKPTRFSTIVVBEHGEGHGLTHFLMNEVIK 130
Qy 135 LOEELNQBKGQKQVLLRCCQLO-Q-EHLGLAETRAEGLHOLEADHSRMKREVAHFHVL 193
Db 131 LQOQMKARDLQRCLELLARLOLEDEKQMTLTVRE-LLTFTQERYKMKREDSYDNLVK 189
Qy 194 LKDEMLSLSHVSNALQKELAAASCRSLQBELYLLKQELQORANWSSCELEQEQSLRT 253
Db 190 VDDNYNLAMYAQLSEKKNVMSRDLQLEIDLQKHLNK--MEECKLE-RNQSLKL 246
Qy 254 ASDQES--GDEELNRLKPEENKLSLFTSL-----AKDILLESLEAR 295
Db 247 KNDIENRPKEQVLELENEMLTKNOELQSIQAGKRSPLSDSKAIDLTLEHDKREAL 306
Qy 296 GSRQLVERIHSRRAVAERQYWEKEQTLLOFKSKMACQLYREKVNALQAVC 355
Db 307 EDQELVNRVNLQEAQAEELRDYLEEKEDLEKCSLTKGKCEMYKHMNTVMLQLE 366
Qy 356 ELQKRDQAYSARDSAQREISQSLVEKDSRLSRQVFLTDQ-----VCELRT 401
Db 367 EVERERDQAFHSRDEAQYQSCLLEKDKYKQIRELEKNDEMVRREACIVNLES 426
Qy 402 QLRLQAE-----PPGVLUK-----EARTPCPCREK----- 428
Db 427 KLRLSKDSNNLDQSLPRNLPTIISODFGDASPTNGQEAADDSSTSESPDSKYFLPY 486
Qy 429 ---QRLVRHAI-CPROSDCSLVSTESQLI-----SDLSATS--SRELVDSPRSSP 476
Db 487 HPQRRNLKGLQLOKASPSILKTSDFQAKGHEEGTDASPSGSLPTNFTSKMP 546
Qy 477 -----APPSQSLYKRVVAEDFGEPEWFSFSCLETPGDPGALPGAKA-GDPHL 523
Db 547 PRSRSIMSITAEPGNDISIVRYKEDAFHR-----STVE-EDNDSGGFDALDDDSHE 600
Qy 524 DY-----ELLDADLPOLLESSLOPVS--PGRLDVSESGVLMRRRPAR 564
Db 601 RYSGFGSSITHSSSSHQSGGLDAYDLEQVNLNMFKESLERFPRSVTSVGHVGRGCP 658
Qy 565 ILSQVTMLAFQGDALLEQISVIGGNLTGIFTHRVTPGSAADQMALRPGTIVMVDYASE 624
Db 659 ----VQHTTLNGDSLTSQTLGGNARGSFVHSPGSLAEGAGLREGHQLLLEGCI 714

Qy 625 PLFKAVLEDTTLEEAVALLRVDGFCCLSVKNTDGYKRLQLDLEAKVATSGDSFYIRVN 684
Db 715 ERQSVPLDCTKEEAHWTIQRCSPVTLHYKVNHEGYRKLKMDMEDGLITSGDSFYIRLN 774
Qy 685 LAMEGRKAG-ELQVHCNEVLHVTDTMFOGCGCWAHHRVNSYTMKDTAAHGTPINYSRAQ 743
Db 775 LNISSQLDACTMSLKCDVHVHVDRTMYODRHEWPCARVDPFDHDLDM-GTIPSYISRAQ 833
Qy 744 QLIALIQDMTOQ-----CTVTRKP-----SSGGPQ-----KLV 771
Db 834 LLVLKQLRMHGRSREEDVGTHTLRLRNTLQPEALSTSDPRVSPRLSRASFLFGOLL 893
Qy 772 RIVSMDKAK-----ASPL-----RLSPD-----RGQLDPSRMESGSCF 805
Db 894 QFVSRSENKRYKRMNSNERVRIISGSPGLSLARSSLDATKLLTEKOEELDPESELGKN--- 950
Qy 806 WAESCLTLPYTLVWPHRPARPPVLLVPRAVGKILSKLCLQG---FKKLAELYSQE 862
Db 951 -----LSLIPYSLVAFYFCERRRPVLTFTVLAKTLVQRLNLSGGAMEFTICKSDIVTRD 1005
Qy 863 EYEAWSORGDIT--QEGEVSGRCWVTRHVESLMEKNTHALLDVLQDLSVCTLHRMDIFP 920
Db 1006 EFLRROKTETIYSREKNPNAFEC-IAPANIEAFAAANKHCLLEAGIGCTRLIKSIY 1064
Qy 921 IVHVSVNEKMAKKUGLQRLGTSEBQLLEAARQEEGLDRAPCLYSLAPDGHSDLDG 980
Db 1065 IVLFRVCEKNIKRFRKLLRPET-EEFLRVCLRKEKELEALPCLYATVEPDMGMSVEE 1123
Qy 981 LLSCVQAATADEQKVVWTEQ 1001
Db 1124 LLRVADKIGEQRKIWYDE 1144

RESULT 4

CARA_HUMAN STANDARD; PRT; 1032 AA.
ID CARA_HUMAN Q9BWT7; Q9UGR5; Q9UGR6; Q9Y3H0;
AC Q9BWT7; Q9UGR5; Q9UGR6; Q9Y3H0;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Caspase recruitment domain protein 10 (CARD-containing MAGUK protein 3) (Carma 3).
DE CARD10 OR CARWA3.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21292987; PubMed=11259443;
RA Wang L., Guo Y., Huang W.-J., Ke X., Poyet J.-L., Manji G.A.,
RA Merriam S., Glucksmann M.A., Distefano P.S., Alnemri E.S., Bertin J.;
RT "CARD10 is a novel caspase recruitment domain/membrane-associated
RT guanylate kinase family member that interacts with Bcl10 and activates
RT NF-kappa B".
RL J. Biol. Chem. 276:21405-21409(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX PubMed=11356195;
RA Gaide O., Martinon F., Mischeau O., Bonnet D., Thome M., Tschopp J.;
RT "Cardinal, a CARD-containing binding partner of Bcl10, induces Bcl10
RT phosphorylation and NF-kappaB activation".
RL FEBS Lett. 496:121-127(2001).
RN [3]
RP ERRATUM.
RA Gaide O., Martinon F., Mischeau O., Bonnet D., Thome M., Tschopp J.;
RL FEBS Lett. 505:198-198(2001).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
RA Clamp M., Smit L.J., Alnscough R., Almeida J.P., Babbage A.,

RA Baggaley C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,
RA Blakey S., Bridgeman A.M., Buck D., Burgess J., Burrill W.D.,
RA Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M.,
RA Cobley V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,
RA Coville G.J., Cox A.V., Davis J., Dawson E., Dhani P.D., Dockree C.,
RA Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,
RA Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E.,
RA Grahm D., Griffiths M.N., Hall C., Hall R., Hall-Tamlyn G.,
RA Heathcote R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J.,
RA Kimberley A., King A., Laird G.K., Langford C.F., Leversha M.A.,
RA Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M.,
RA Matthews L., Mccann O.T., Mccall J., Mclaren S., Mcurray A.A.,
RA Milne S.A., Mortimore B.J., Odeh C.N., Pavitt R., Pearce A.V.,
RA Pearson D., Phillimore B.J., Phillips S.H., Plumb R.W., Ramsay H.,
RA Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D.,
RA Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,
RA Sulston J.E., Swann R.M., Vaudin M., Walli M., Wallis J.M.,
RA Whiteley M.N., Willey D., Williams L., Williams S., Williamson H.,
RA Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,
RA Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,
RA Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,
RA Mitsuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S.,
RA Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S.,
RA Lai H., Lao H.I., Lewis J., Lewis S., Lin S.-P., Loh P., Mala J.E.,
RA Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S.,
RA Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D.,
RA Wu H., Yao Z., Zhan M., Zhang G., Chissoe S., Murray J., Miller N.,
RA Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H.,
RA Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T.,
RA Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P.,
RA Rohlfing T., Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K.,
RA Nelson J., Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R.,
RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA Edelmann L., Kim U.J., Shizuwa H., Simon M.I., Dumanski J.P.,
RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA Khan A.S., Lane L., Tilahun Y., Wright H.,
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
CC -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK.
CC -!- SUBUNIT: CARD10 and Bcl10 bind to each other by CARD-CARD
CC interaction. They both participate in a complex with MALTL1, where
CC MALTL1 binds to Bcl10 (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- TISSUE SPECIFICITY: Detected in adult heart, kidney and liver;
CC lower levels in intestine, placenta, muscle and lung. Also found
CC in fetal lung, liver and kidney.
CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -!- CAUTION: Supposed to contain a SH3, a PDZ and a guanylate kinase-
CC like domain. But none of these 3 domains are detected by PROSITE,
CC Pfam or SMART.
CC -!- CAUTION: Ref.4 sequence differs from that shown due to various
CC gene identification problems.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS0209; CARD; 1.
CC Coiled coil.
CC -----
CC DR EMBL; AY028896; AAK26165.1; -;
CC DR EMBL; AY032928; AAK54454.1; -;
CC DR EMBL; AL0498

| RESULT | CARD9_RAT | CARD9_HUMAN | STANDARD | PRT | 536 AA |
|--------|--|-------------|--|-----|--------|
| 283 | LEEDWRQALQEHQEQASTIFSRLKDLRQAEALRTRCMEEKEMFELQCLALRKDKAKWKDR | 347 | VNALQAVCELOKQEQARDQAYSAQSAQREISQSLVEKSLSRQVFEITDQVCELRQLQRL | 406 | |
| 284 | QEPYQ | 348 | QEPYQ | 407 | |
| 285 | 01-MAR-2002 (Rel. 41, Created) | 349 | 01-MAR-2002 (Rel. 41, Last sequence update) | 408 | |
| 286 | 01-MAR-2002 (Rel. 41, Last sequence update) | 350 | 01-MAR-2002 (Rel. 41, Last sequence update) | 409 | |
| 287 | 01-MAR-2002 (Rel. 41, Last annotation update) | 351 | 01-MAR-2002 (Rel. 41, Last annotation update) | 410 | |
| 288 | Caspase recruitment domain protein 9 (rCARD9) | 352 | Caspase recruitment domain protein 9 (hCARD9) | 411 | |
| 289 | CARD9 | 353 | CARD9 | 412 | |
| 290 | Rattus norvegicus (Rat) | 354 | Rattus norvegicus (Rat) | 413 | |
| 291 | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | 355 | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | 414 | |
| 292 | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. | 356 | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. | 415 | |
| 293 | NCBI_TaxID=10116; | 357 | NCBI_TaxID=10116; | 416 | |
| 294 | SEQUENCE FROM N.A. | 358 | SEQUENCE FROM N.A. | 417 | |
| 295 | STRAIN=Sprague-Dawley; | 359 | STRAIN=Sprague-Dawley; | 418 | |
| 296 | MEDLINE=20576268; PubMed=11053425; | 360 | MEDLINE=20576268; PubMed=11053425; | 419 | |
| 297 | Bertin J., Guo Y., Wang L., Srinivasula S.M., Jacobson M.D., | 361 | Bertin J., Guo Y., Wang L., Srinivasula S.M., Jacobson M.D., | 420 | |
| 298 | Poyet J.-L., Merriam S., Du M.-Q., Dyer M.J.S., Robison K.E., | 362 | Poyet J.-L., Merriam S., Du M.-Q., Dyer M.J.S., Robison K.E., | 421 | |
| 299 | Distefano P.S., Alnemri E.S.; | 363 | Distefano P.S., Alnemri E.S.; | 422 | |
| 300 | "CARD9 is a novel caspase recruitment domain-containing protein that | 364 | "CARD9 is a novel caspase recruitment domain-containing protein that | 423 | |
| 301 | interacts with Bcl10/CLAP and activates NF-kappa B."; | 365 | interacts with Bcl10/CLAP and activates NF-kappa B."; | 424 | |
| 302 | J. Biol. Chem. 275:41082-41086(2000). | 366 | J. Biol. Chem. 275:41082-41086(2000). | 425 | |
| 303 | -I- FUNCTION: Activates NF-kappa B via Bcl10 (By similarity). | 367 | -I- FUNCTION: Activates NF-kappa B via Bcl10 (By similarity). | 426 | |
| 304 | -I- SUBUNIT: Self-associates. CARD9 and Bcl10 bind to each other by | 368 | -I- SUBUNIT: Self-associates. CARD9 and Bcl10 bind to each other by | 427 | |
| 305 | CARD-CARD interaction (By similarity). | 369 | CARD-CARD interaction (By similarity). | 428 | |
| 306 | -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity). | 370 | -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity). | 429 | |
| 307 | -I- SIMILARITY: CONTAINS 1 CARD DOMAIN. | 371 | -I- SIMILARITY: CONTAINS 1 CARD DOMAIN. | 430 | |
| 308 | ----- | 372 | ----- | 431 | |
| 309 | This SWISS-PROT entry is copyright. It is produced through a collaboration | 373 | This SWISS-PROT entry is copyright. It is produced through a collaboration | 432 | |
| 310 | between the Swiss Institute of Bioinformatics and the EMBL Outstation - | 374 | between the Swiss Institute of Bioinformatics and the EMBL Outstation - | 433 | |
| 311 | the European Bioinformatics Institute. There are no restrictions on its | 375 | the European Bioinformatics Institute. There are no restrictions on its | 434 | |
| 312 | use by non-profit institutions as long as its content is in no way | 376 | use by non-profit institutions as long as its content is in no way | 435 | |
| 313 | modified and this statement is not removed. Usage by and for commercial | 377 | modified and this statement is not removed. Usage by and for commercial | 436 | |
| 314 | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | 378 | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | 437 | |
| 315 | or send an email to license@isb-sib.ch). | 379 | or send an email to license@isb-sib.ch). | 438 | |
| 316 | ----- | 380 | ----- | 439 | |
| 317 | EMBL; AF311288; AAC28791.1; | 381 | EMBL; AF311288; AAC28791.1; | 440 | |
| 318 | InterPro; IPR001315; CARD. | 382 | InterPro; IPR001315; CARD. | 441 | |
| 319 | PROSITE; PS50209; CARD; 1. | 383 | PROSITE; PS50209; CARD; 1. | 442 | |
| 320 | Coiled coil. 6 98 | 384 | Coiled coil. 6 98 | 443 | |
| 321 | DOMAIN | 385 | DOMAIN | 444 | |
| 322 | FT DOMAIN 117 277 | 386 | FT DOMAIN 117 277 | 445 | |
| 323 | FT DOMAIN 303 420 | 387 | FT DOMAIN 303 420 | 446 | |
| 324 | SEQUENCE 536 AA; 62631 MW; 6F33089CB7E6BAC9 CRC64; | 388 | SEQUENCE 536 AA; 62631 MW; 6F33089CB7E6BAC9 CRC64; | 447 | |
| 325 | Query Match | 389 | Query Match | 448 | |
| 326 | Best Local Similarity 10.08; Score 512.5; DB 1; Length 536; | 390 | Best Local Similarity 10.08; Score 512.5; DB 1; Length 536; | 449 | |
| 327 | Matches 162; Conservative 94; Mismatches 215; Indels 91; Gaps 15; | 391 | Matches 162; Conservative 94; Mismatches 215; Indels 91; Gaps 15; | 450 | |
| 328 | 15 DEPTLWEMSHRRIVRCIPCSRLTPYLRQAKVLCQDEEVLHSPRLTNSAMRAGHLL 74 | 392 | 15 DEPTLWEMSHRRIVRCIPCSRLTPYLRQAKVLCQDEEVLHSPRLTNSAMRAGHLL 74 | 451 | |
| 329 | 6 NDDECNALSALESFRVRLISVIDPSRITPYLRCKVLNPDDEQVLSDPNLVIRKRVGVLL 65 | 393 | 6 NDDECNALSALESFRVRLISVIDPSRITPYLRCKVLNPDDEQVLSDPNLVIRKRVGVLL 65 | 452 | |
| 330 | 75 DLLKTRKNGAIAFLSLKFNPDVYTLVTGLQPD-----VDFSNFSL-----ME 120 | 394 | 75 DLLKTRKNGAIAFLSLKFNPDVYTLVTGLQPD-----VDFSNFSL-----ME 120 | 453 | |
| 331 | 66 DILQRTKHGKGVAFLESLYLYPOLYRKVTGKEPARVFESMIIDAGSGSLTQLLMTVMK 125 | 395 | 66 DILQRTKHGKGVAFLESLYLYPOLYRKVTGKEPARVFESMIIDAGSGSLTQLLMTVMK 125 | 454 | |
| 332 | 121 TSKLTECLAGNIGLSQELNQEKGKQVLLRRCQQLQEHGLAETRAEGLHQLADHSM 180 | 396 | 121 TSKLTECLAGNIGLSQELNQEKGKQVLLRRCQQLQEHGLAETRAEGLHQLADHSM 180 | 455 | |
| 333 | 126 LQKVKQDLTALLSS-KDQFIKELRVKDSLLRKHQVQ-----RL | | | | |

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AF311287; AAG28790.1; -;
CC DR EMBL; AK024001; BAB14766.1; ALT_FRAME.
CC DR EMBL; BC008877; AAH08877.1; -;
CC DR InterPro: IPR001315; CARD.
CC DR PROSITE; PS50209; CARD; 1.
KW Coiled coil.
FT DOMAIN 6 98 CARD.
FT DOMAIN 117 277 COILED COIL (POTENTIAL).
FT DOMAIN 332 419 COILED COIL (POTENTIAL).
FT CONFLICT 12 12 N -> S (IN REF. 3).
FT CONFLICT 482 482 LSSGEPPEKER -> PAGLPGIGAVC (IN REF. 3).
SQ SEQUENCE 536 AA; 62267 MW; 6EB1835315B83DE5 CRC64;

Query Match 9.6%; Score 493; DB 1; Length 536;
Best Local Similarity 29.5%; Pred. No. 6.8e-18;
Matches 155; Conservative 92; Mismatches 200; Indels 78; Gaps 14;

QY 15 DEETLWEMESHRRIVRCIPSLPYLRQAKVLQQLDEEVLHSPRLTNSAMRAGHLL 74
Db 6 NDDECNWVLEGFRTLTSDVIDPSRITPYLRQCKVLNPDDEEQVLSDPNLVIRKRVGVLL 65
QY 75 DILKTRGKNGAIFLESFKFHNPDVTLVTLGLOPD-----VDFSNFSGL-----ME 120
Db 66 DILQRTGHGKGVAFLESLEYLPOLYKVTGKEPARVFSMIIDASGSLTQLLMTVMK 125
QY 121 TSKLTCLAGAGISLOELNQEKGQEVLLRRCQQLQEHGLGLAETRAEGLHQLLEADHSM 180
Db 126 LQKKVQDLTALLSS-RDDFKELRVKDSLRLKHQVQ-----RL 164
QY 181 KREVSAHFHVLEVKDEMLSLHYNAQKEKELASRCRSQELLYLLKQELQORANMWS 240
Db 165 KEECEAGSRELKKEENYDLAMRLAQHSEKGAALNRNDLQLEIDQLKHSMLKAE--D 222
QY 241 SCELEQEQ-SLTASDOESGDEELNRLKEE---NEKLSLTFSLAEK-----DI 286
Db 223 DCKVERKHUKLRHAEQRPSEQLLEWELQEKALLQARVOELASVOEGKLDSSPYIQV 282
QY 287 LEQSLDEARGSGRELVERIHSRLRAVAERQEQWEEKQETLLQFQSKMACQLYREK 346
Db 283 LEEDWRQALRDHOEQANTIFSLRKDLRQGEARRLCRMEKEFELQCLALRKDSKMYKDR 342
QY 347 VNALQAVCELOKEROQVYARDSAORETSQSILVEKDSLRQVFEVLTQVCELTQLRQL 406
Db 343 IEAILLOMEEVAIERDQAIATREELHAQHARGLOEKDALRKQVRELGEKADELQQLVQC 402
QY 407 QAEPPGVLEKQAEATREPCPREKQRLVRMHAI---CPRDDSDCSILVSTESQLLSD---- 458
Db 403 EAQ---LLAVEGLRR---QQLFTVLSSDLEDGSPRRSOELSLPQDLEDTLQSDKGLA 456
QY 459 -----LSATSRRELVDSEFRS---SSPAPPQQSLYKRAEDF 492
Db 457 GGGSPQPFPAALHQEQLRPNRHDAGLSSGEPPEKER--RLKESF 499

RESULT 8

Z01_MOUSE STANDARD; PRT; 1745 AA.
ID Z01_MOUSE
AC P39447;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tight junction protein ZO-1 (Zonula occludens 1 protein) (zona occludens 1 protein) (Tight junction protein 1).
GN TJP1 OR Z01.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=93252986; PubMed=8486731;
RA Itoh M., Nagafuchi A., Yonemura S., Yasuda-Kitani T., Tsukita S., Tsukita S.;
RT "The 220-kD protein colocalizing with cadherins in non-epithelial cells is identical to ZO-1, a tight junction-associated protein in epithelial cells: cDNA cloning and immunoelectron microscopy.";
RL J. Cell Biol. 121:491-502(1993).
CC -!- FUNCTION: THE N-TERMINAL MAY BE INVOLVED IN TRANSDUCING A SIGNAL REQUIRED FOR TIGHT JUNCTION ASSEMBLY, WHILE THE C-TERMINAL MAY HAVE SPECIFIC PROPERTIES OF TIGHT JUNCTIONS. THE ALPHA DOMAIN MIGHT BE INVOLVED IN STABILIZING JUNCTIONS.
CC -!- SUBUNIT: INTERACTS WITH OCCUDIN, CLAUDINS AND ZO-3.
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE. CYTOPLASMIC SIDE. MOVEMENT OF ZO-1 FROM THE CYTOPLASM TO MEMBRANE IS AN EARLY EVENT OCCURRING CONCURRENTLY WITH CELL-CELL CONTACT (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; D14340; BAA03274.1; -;
DR PIR; A6431; A6431.
DR HSSP; P31016; IBE9.
DR MGD; MGI:98759; Tjpl.
DR InterPro: IPR000619; Guanylate_kin.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001452; SH3.
DR InterPro: IPR000906; ZU5.
DR Pfam; PF00825; Guanylate_kin; 1.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00791; ZU5; 1.
DR SMART; SM00072; GUKC; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR SMART; SM00218; ZU5; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 3.
DR PROSITE; PS50002; SH3; 1.
KW Tight junction; SH3 domain; Repeat; Membrane.
FT DOMAIN 23 110 PDZ 1.
FT DOMAIN 186 264 PDZ 2.
FT DOMAIN 421 502 PDZ 3.
FT DOMAIN 516 584 SH3.
FT DOMAIN 644 794 GUANYLATE KINASE.
FT DOMAIN 1242 1247 POLY-PRO.
FT DOMAIN 1424 1430 POLY-PRO.
SQ SEQUENCE 1745 AA; 194710 MW; C3DA2C0A9F411F66 CRC64;

Query Match 5.6%; Score 289; DB 1; Length 1745;
Best Local Similarity 22.3%; Pred. No. 4.5e-07;
Matches 191; Conservative 136; Mismatches 319; Indels 212; Gaps 43;
QY 231 QELQRANWVSSCELEQSLRTASDOESGDEELNRLKEENKLSLTFSLAEKIDLEQS 290
Db 70 QENDRVAMVNGVSNVDEHFAVQQLRKSKNA--KITIRKKKQVIVSHPDPEPVSND 127
QY 291 LDEARGSRQELVERIHSRLRE-RAVAARQREQWEEKQETLLQFQSKMACQLYREKVA 349
Db 128 EDDSYD-----EVHDPFRAGALANRRSEKSWARDSA---SRERSLSPRSDRRSVAS 178

QY 350 LQ-----AQVCELOKQERDQAYARDSAQ---REISQ-SLVEKDSLRQ---VFELTDQVCE 398
 Db 179 SQAPKTKVTLVKSKEEYGLRPASHIFVKEISQSLAARDQIQEGDGVVLKINGVTYE 238
 QY 399 --LRTQLRQLOQAPPGVGLKQEARTRPCPREKQRLVR-----MHAICPRDSDSCSLVS 449
 Db 239 NMSLTDAKTLIERSKGLKLMVQR-----DERATLLNVPDLSIHSANASERDDISEIQ 293
 QY 450 STESQLLSLSATS-----SRELVDSSFRSSPAPPSQOQSLYKRVAEDFGE 495
 Db 294 S-----LASDHSGRSHDRPPRSOSRSPDQSEPSDSTQSPQPSNGSLRSR-EEERMSK 348
 QY 496 PWSFSCLE-----IPSGDGCALPKAGADPHLDYELLDTA 531
 Db 349 PGAISTPVKHVDHPPKAVEEVTVEKNEKOTPLPEPKP---VYAGVQGPVDVL----- 399
 QY 532 DLPQLESSLPVSP--GRLDVS--ESGLMRRRPPARRILLSOVTMLAFQGDALLEQISVIG 587
 Db 400 -----PVSPSGALPNSAHEDGIL---RPSMKLVK-----FRKGDV--GLRLAG 439
 QY 588 GNLGTGIFIHRTPGSAADOMALRPGTQIVMVDYEASEPLFKAVLEDTTL-----EEA 639
 Db 440 GNDVGIFVAGVLEDSFAKREGLQDILLRVN---NVDFNIIIREAVLFLDLDPKGEV 496
 QY 640 VGLLRVDGFCCLSVKVNVDGYKRLQDLLEAKVATSGDSFYIRVNLAMEGRAKGEIQVHC 699
 Db 497 TILAQK-----KKDVYRRIE-----SDVGDSEFYIRTHFEYKESPYGLSFNK 539
 QY 700 NEVLHVTDMFQO--CGCWAHHRVNSYTMKDTAAHGHTIPNYSRAOQQLIALIQDWTQOCTV 758
 Db 540 GEYFRVVDLYNGKLSWLAIRGKNHKE--VERGIIPKNRAEQ--LASVQ-----Y 588
 QY 759 TRKPSGGPKLVRIYMDKAKASPLR--LSFDRGQDLPSPMEGSSCTCFWAESECLTLVP-Y 816
 Db 589 TLPTAGG-----DRADFWREGLRSSRKNLAKSREDLS-----AQPVTTFPAY 633
 QY 817 TLMVHRPARPRVLLVPRVAVGKILSEKLLQGLQFKKLAELYSQEE---YE-AWSQGD 872
 Db 634 ERVVLREAGLRPTIF-GPIADVAREK-----LAREEPDIYQIAKSELRD 678
 QY 873 IIOGEVSGRCWVTHAVESLMEKNTHALLDVQDSVCTLHRMDIFPIVIHVSVEKMA 932
 Db 679 AGTDHRSRG---IIRLHTKIQIIDQKHALLDVTNPNAVDRLNTYAQWYPIVVFVLPNSQKG 735
 QY 933 KKL-----KKGLRQLRQTSQEQLLEAARQEGDIDRAPCLYSSILAPDGSWDLGLLS 983
 Db 736 VKTMRMLCPESRKSARKLYERSHKL-----RKNHHLFTTINLSN-NDGW-----YG 784
 QY 984 CVRQAIADQKVVWTEQ 1001
 Db 785 ALKEAIOQQOQNLVWVSE 802

RESULT 9
 Z02_MOUSE
 ID Z02_MOUSE STANDARD; PRT; 1167 AA.
 AC Q920U1;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Tight junction protein ZO-2 (Zonula occludens 2 protein) (Zona
 DE occludens 2 protein) (Tight junction protein 2).
 GN TJP2 OR ZO2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Chordata; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C56BL/6 X CBA;
 RX MEDLINE=99150392; PubMed=10026224;
 RA Itoh M., Morita K., Tsukita S.;
 RT "Characterization of ZO-2 as a MAGUK family member associated with

RT tight as well as adherens junctions with a binding affinity to
 RT occludin and alpha catenin.";
 RL J. Biol. Chem. 274:5981-5986(1999).
 CC -!- FUNCTION: PLAYS A ROLE IN TIGHT JUNCTIONS AND ADHERENS JUNCTIONS.
 CC -!- SUBUNIT: INTERACTS WITH OCCLUDIN, MEMBRANE, CYTOPLASMIC SIDE.
 CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE.
 CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF113005; AAD19964.1; -
 DR HSSP: P31016; 1BE9.
 DR MGD: MGI:1341872; Tjp2.
 DR InterPro: IPR000619; Guanylate_kin.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00625; Guanylate_kin; 1.
 DR Pfam: PF00595; PDZ; 3.
 DR Pfam: PF00018; SH3; 1.
 DR SMART: SM00072; GUKc; 1.
 DR SMART: SM00228; PDZ; 3.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
 DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
 DR PROSITE: PS50106; PDZ; 3.
 DR PROSITE: PS50002; SH3; 1.
 KW Tight junction; SH3 domain; Repeat; Membrane.
 FT DOMAIN 10 97
 FT DOMAIN 287 365
 FT DOMAIN 489 570
 FT DOMAIN 584 649
 FT DOMAIN 678 858
 FT DOMAIN 1139 1142
 FT POLY-GLU.
 SQ SEQUENCE 1167 AA; 131614 MW; F15DA3EBC3F9434F CRC64;

Query Match 5.4%; Score 280; DB 1; Length 1167;
 Best Local Similarity 23.0%; Pred. No. 7.7e-07;
 Matches 198; Conservative 122; Mismatches 308; Indels 234; Gaps 43;
 QY 248 EQSLRTASDQESGDEELNR-----LKEENKLSRLTFLAEKDI----- 286
 Db 134 ERSRSHSHDMLSHSWEGRERGGPHQTSRERSRGRSLERGLDQEDYGRSRSRGR 193
 QY 287 -LEQSLDEARGSQRELVERIHSRERAVAAERQR-----EQY-----WEEK 327
 Db 194 SLERGLD-----RDFVSRDHS--RGRSIDRDYDRYSYHEAYEPDYGGSYSYDRA 246
 QY 328 QTLQFOKSKMACOLYREKYNALQAO-----VCELOKQERDQAYSARSDA 371
 Db 247 HPETRYERSRREHL-RSRSPSPESRHEHKGHDPRPGVLLTOSKANEETGLRGS 305
 QY 372 Q---REISQS-LVEKDSLRQ---VFELTDQVCE--LRTQLRQLOQAPPGVGLKQEARTR 422
 Db 306 QIFIKEMTRCTLATKGNLHEGDIILKINGVTNMSLTDLKLEKSRGKL-QLVLRD 364
 QY 423 PCPREKQRLVNMHAICPRDD--SDCLSVSTES-----QLLSDLSATSRRELVDSPRS 473
 Db 365 ----SKQTLINIPALNDRDSEVEDISETSNRSPSPERRQOYSDQDYSTSEKL----- 415
 QY 474 SSPAPPSSQOQLYKRVAEDFGEPPWSFSSCLEI-PEGDPGALPGAKAGDPHLDYELLDTA 532
 Db 416 -KERPSSRETSRGLSR-MGATPTPFKTDITPAGVTEA-----TREPRYQEE-----GP 464

```
QY 533 LPQLESSLPVSPGRDVSSEGVLMRRRPPARRIL--SQVTLMAF-QGDALLEQISVTGGN 589
Db 465 VP-----QRAAPR-----VFLRSPDEAIYGNTRKVRFKGDSV--GLRLPGGN 509
QY 590 LTGFIHRVTPGSAADOMLRPGQIIVMDYEAEPFLFKAVLEDTTL-----EEAVGL 642
Db 510 DVGIFVAGIPEGTSRAGQEGQIILKVNFDPRGL---VREDAVLYLLEIPKGETVTI 566
QY 643 LRRVDGFCCLSVKNTDGYKRLLOLEAKVATSGDSFYIRVNLAMEGRAGKELGVHNEV 702
Db 567 L-----AQSRADVYRDIL-----ACGRGDSFFIRSHFCEKETPQSLAFTRGV 610
QY 703 LHVTDTFQCGCGWHAHRVNSYTMKDTAAHTTTPNYSRAQQOQIALIQMTQCT---- 757
Db 611 FRVYDTLYDGKGLHMLAVRIGNDLK-----GLIPNKSRAEQ-----MDSVQNAQRENA 659
QY 758 -----VTRKPSGGPQKLVIRVSMKAKASPLRLSFGOLDPSRMESGSCFCFAES 809
Db 660 GDAVFWMMRRQSRGGDKTLR-----KSRDLARSYSVSTKFFA-- 700
QY 810 CLTVLYPTLVMPHRRPARRPVLL-----VPRAVGKILSEKLCGLQFKKCLAEYLSQBEYE 865
Db 701 -----YEKVLRLREAGFKRPVLEGPDIADIAERLANELPDLFOTAK-----TEPKD 746
QY 866 AWSORGDIQEGEYSGRCWVTRHAVESLMEKNTHALLDVQDSVCTLHRMDIPIPIHV 925
Db 747 AGSEKSGV-----VRLATVQIIEQDKHALLDVTPKRAVDLLNTQWPIVIF 795
QY 926 SVNEKMAKLLKGLQLR-----GTSEQLLEAARQEGDLDR--APCLYSSILAPDGSGLD 979
Db 796 NPDSRQGVKTR--QRLSPTSNKSRLKLDQANKLKTCSHLFTATINVSANDGW---- 849
QY 980 GLLSCVQAIADEQKVVWTEQ 1001
Db 850 --FGLSKDSIQOQNEAVWVSE 869

RESULT 10
ZOL_HUMAN
ID_ZOI_HUMAN STANDARD; PRT: 1736 AA.
AC Q07157;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tight junction protein ZO-1 (zonula occludens 1 protein) (Zona
DE occludens 1 protein) (Tight junction protein 1).
GN TJP1 OR ZO1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93361541; PubMed=8395056;
RA Willott E., Balda M.S., Fanning A.S., Jameson B., van Itallie C.,
RA Anderson J.M.;
RT "The tight junction protein ZO-1 is homologous to the Drosophila
RT discs-large tumor suppressor protein of septate junctions.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:7834-7838(1993).
CC CC -1- FUNCTION: THE N-TERMINAL MAY BE INVOLVED IN TRANSDUCING A SIGNAL
CC REQUIRED FOR TIGHT JUNCTION ASSEMBLY, WHILE THE C-TERMINAL MAY
CC HAVE SPECIFIC PROPERTIES OF TIGHT JUNCTIONS. THE ALPHA DOMAIN
CC BEING INVOLVED IN STABILIZING JUNCTIONS.
CC -1- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-3.
CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE. CYTOPLASMIC SIDE.
CC MOVEMENT OF ZO-1 FROM THE CYTOPLASM TO MEMBRANE IS AN EARLY EVENT
CC OCCURRING CONCURRENTLY WITH CELL-CELL CONTACT.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: THE ALPHA-CONTAINING ISOFORM IS FOUND IN MOST
CC EPITHELIAL CELL JUNCTIONS. THE SHORT ISOFORM IS FOUND BOTH IN
CC ENDOTHELIAL CELLS AND THE HIGHLY SPECIALIZED EPITHELIAL JUNCTIONS
```

```
CC OF RENAL GLOMERULI AND SERTOLI CELLS OF THE SEMINIFEROUS TUBULES.
CC -1- PTM: PHOSPHORYLATED.
CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: L14837; AAA02891.1; -.
CC PIR: A47747; A47747.
CC HSP: P31016; 1BFE.
CC MIM: 601009; -.
CC InterPro: IPR000619; Guanylate_kin.
CC InterPro: IPR001478; PDZ.
CC InterPro: IPR001452; SH3.
CC InterPro: IPR000906; ZU5.
CC Pfam: PF00625; Guanylate_kin; 1.
CC Pfam: PF00595; PDZ; 3.
CC Pfam: PF00018; SH3; 1.
CC Pfam: PF00791; ZU5; 1.
CC SMART: SM00072; GUKC; 1.
CC SMART: SM00228; PDZ; 3.
CC SMART: SM00326; SH3; 1.
CC SMART: SM00218; ZU5; 1.
CC PROSITE: PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
CC PROSITE: PS00052; GUANYLATE_KINASE_2; 1.
CC PROSITE: PS0106; PDZ; 3.
CC PROSITE: PS00002; SH3; 1.
CC Tight junction; SH3 domain; Alternative splicing; Repeat; Membrane;
CC Phosphorylation.
CC DOMAIN 11 98 PDZ 1.
CC DOMAIN 174 252 PDZ 2.
CC DOMAIN 409 490 PDZ 3.
CC DOMAIN 504 572 SH3.
CC DOMAIN 632 782 GUANYLATE KINASE.
CC DOMAIN 1231 1236 POLY-PRO.
CC DOMAIN 1414 1420 POLY-PRO.
CC VARSPIC 910 989 MISSING (IN SHORT ISOFORM).
CC SEQUENCE 1736 AA; 194721 MW; 508D01B7A0814FFE CRC64;

Query Match 5.3%; Score 270.5; DB 1; Length 1736;
Best Local Similarity 21.4%; Pred No. 3.8e-06;
Matches 174; Conservative 129; Mismatches 278; Indels 233; Gaps 38;

QY 247 QEQSLRTASDQESGDEE-----LNRLKEENKLRSLTFSLAPEKDLQSLDEARG 296
Db 151 RERSLSPRRSRSVASSQAPKPTKVLVKSRSKNEYGLR-LASHIFVKESIQSDSLAARDG 209
QY 297 SRQE--LVERIH-SLRERAVAAERQYWEKEQTLLQFOKSKMACQLTREKYNALQAO 353
Db 210 NIOEGDVVLKINGTVTNMSLTD-----AKTILERSKGLKMWVQRDE----RAT 255
QY 354 VCELOKRDQAYSARDSAOREIS--QSLVEKDSLRQVFEITDQVCELRTLRLQLQAEPP 411
Db 256 LLNVPDLSDSIHSANASERDDISQIQLASHDSGERS-----HRRPP 296
QY 412 GVLKQEAARTPCPRKQRLVRMHAICPRDDSDGSLVSTTESQL-----LSDL 459
Db 297 ----RRSRSPDQSRPSDHSRHS--PQPSNGSLRSRDEERISKPGAVSTPVKHADDH 350
QY 460 SATSREL-VDSFRSSSPAPPQQSLYKRVAEDEGCEPWFSSCLEIPEGDPGALPAGA 518
Db 351 TPKTVEEVTVERNKQTPSLPEPKPVAAQV---GNQMWYLSV----- 390
QY 519 GDPHLDYELLDTADLPQLESSLPVSPGRDVSSEGVLMRRRPPARRILSQVTLMAFQGA 578
```


Db 559 --LLEIPKGEWVTLAQRADVYRDL-----ACGRGDSFFIRSHFECEKETPQSIAFS 610
 Qy 699 CNEVLHVTDTMFOG-CGCWHAHRVNSYTMKDAAHGTIPNYSRAQQOIALIQDMTOOCT 757
 Db 611 RGEVFRVDTLYDGLKHLWLAVRIGNELEK-----GLIPNKSRAEQ--MASVQN----- 657
 Qy 758 VTRKPSGGQPKLVRIYSMDKAKAPLSPLSDRQQLDPSRMWEGSTCFWABSCULVLPYT 817
 Db 658 -AQDNAGDRADFRWM-----RGQSGMKNLKRSREDLTAASVSTKPPA-----YE 704
 Qy 818 LVNPHRPARPVPVLL-----VPRAVGKILSEKLCLOQFKKLAEYLSQEEVEAWSORGDI 873
 Db 705 RVLLREAGFRPVVFGPIADIALEKLANELPDFQAK-----TEPKDAGESSG 756
 Qy 874 IQEVEVGGRCWTRHAVESLMKNTHALDVQDLSVCTLRHMDIFFPIVIRVSNKMAK 933
 Db 757 V-----VRLNTRVQIIQEDKHALLDVTPKAVDLLNNTQWFFIVIFFNPDSDRGV 805
 Qy 934 KKKGLQRLGTSEBQLLEAARQEGDLDRAPLY-----SSLAPGWSDDLGLLSCVR 986
 Db 806 KTRM--ORLNPSTSNKSRKLYDQANKLKKT-CAHLETFATINLNSANDSW-----FGSLK 856
 Qy 987 QAIADQKQVWVTEQ 1001
 Db 857 DTIQHQGGEAVWVSE 871

RESULT 12
 ZO2_HUMAN
 ID ZO2_HUMAN STANDARD; PRT: 1190 AA.
 AC Q9UDY2; Q15883; Q9UDY1; Q9UDY0; Q99839;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Tight junction protein ZO-2 (Zonula occludens 2 protein) (zona
 occludens 2 protein) (Tight junction protein 2).
 GN TJP2 OR ZO2 OR X104.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 [1]
 RN SEQUENCE FROM N.A. (ISOFORM A1).
 RC TISSUE=Brain;
 RX MEDLINE=95038744; PubMed=7951235;
 RA Duclos F., Rodius F., Wrognemann K., Mandel J.L., Koenig M.;
 RT "The Friedreich ataxia region: characterization of two novel genes and
 RL reduction of the critical region to 300 kb.";
 RN Hum. Mol. Genet. 3:909-914(1994).
 [2]
 RN SEQUENCE FROM N.A. (ISOFORMS A1; C1; A2 AND C2).
 RC TISSUE=Pancreas;
 RX MEDLINE=20472048; PubMed=11018256;
 RA Chlenski A., Ketels K.V., Korovaitseva G.I., Talamonti M.S., Oyasu R.,
 RA Scarpelli D.G.;
 RT "Organization and expression of the human zo-2 gene (tjp-2) in normal
 RT and neoplastic tissues.";
 RL Biochim. Biophys. Acta 1493:319-324(2000).
 [3]
 RN SEQUENCE OF 1-104 FROM N.A. (ISOFORMS A1 AND C1).
 RC TISSUE=Pancreas;
 RX MEDLINE=99287578; PubMed=10360833;
 RA Chlenski A., Ketels K.V., Tsao M.-S., Talamonti M.S., Anderson M.R.,
 RA Oyasu R., Scarpelli D.G.;
 RT "Tight junction protein ZO-2 is differentially expressed in normal
 RT pancreatic ducts compared to human pancreatic adenocarcinoma.";
 RL Int. J. Cancer 82:137-144(1999).
 [4]
 RN PARTIAL SEQUENCE FROM N.A. (ISOFORMS A1 AND C1).
 RC TISSUE=Pancreas;
 RX MEDLINE=99426875; PubMed=10495427;
 RA Chlenski A., Ketels K.V., Engeriser J.L., Talamonti M.S., Tsao M.-S.,
 RA Koutnikova H., Oyasu R., Scarpelli D.G.;

RT "Zo-2 gene alternative promoters in normal and neoplastic human
 RT pancreatic duct cells.";
 RL Int. J. Cancer 83:349-358(1999).
 RN [5]
 RP SEQUENCE OF 1047-1167 FROM N.A.
 RC TISSUE=Aortic smooth muscle;
 RA Adams L.D., Werny I., Schwartz S.M.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PLAYS A ROLE IN TIGHT JUNCTIONS AND ADHERENS JUNCTIONS.
 CC -1- SUBUNIT: INTERACTS WITH OCCLUDIN.
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE. CYTOPLASMIC SIDE (BY
 CC SIMILARITY).
 CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; A1 (SHOWN HERE), C1, A2 AND C2;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING AND ALTERNATIVE INITIATION.
 CC -1- TISSUE SPECIFICITY: THIS PROTEIN IS FOUND IN EPITHELIAL CELL
 CC JUNCTIONS. ISOFORM A1 IS ABUNDANT IN THE HEART AND BRAIN WHEREAS
 CC ISOFORM C1 IS EXPRESSED AT HIGH LEVEL IN THE KIDNEY, PANCREAS,
 CC HEART AND PLACENTA. IN BRAIN AND SKELETAL MUSCLE, ONLY ISOFORM A1
 CC IS DETECTABLE. ISOFORM C1 IS FOUND IN NORMAL AS WELL AS IN MOST
 CC NEOPLASTIC TISSUES WHILE ISOFORM A1 IS PRESENT ALMOST EXCLUSIVELY
 CC IN NORMAL TISSUE.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITION 1086
 CC ONWARD DUE TO A FRAMESHIFT.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch)
 CC -----

DR EMBL; L27476; AAA61300.1; ALT_FRAME.
 DR EMBL; L27476; AAA61300.1; ALT_FRAME.
 DR EMBL; AF177533; AAD20387.2; JOINED.
 DR EMBL; AF043195; AAD20387.2; JOINED.
 DR EMBL; AF043196; AAD20387.2; JOINED.
 DR EMBL; AF043197; AAD20387.2; JOINED.
 DR EMBL; AF177518; AAD20387.2; JOINED.
 DR EMBL; AF177519; AAD20387.2; JOINED.
 DR EMBL; AF177520; AAD20387.2; JOINED.
 DR EMBL; AF177521; AAD20387.2; JOINED.
 DR EMBL; AF177522; AAD20387.2; JOINED.
 DR EMBL; AF177523; AAD20387.2; JOINED.
 DR EMBL; AF177524; AAD20387.2; JOINED.
 DR EMBL; AF177525; AAD20387.2; JOINED.
 DR EMBL; AF177526; AAD20387.2; JOINED.
 DR EMBL; AF177527; AAD20387.2; JOINED.
 DR EMBL; AF177528; AAD20387.2; JOINED.
 DR EMBL; AF177529; AAD20387.2; JOINED.
 DR EMBL; AF177530; AAD20387.2; JOINED.
 DR EMBL; AF177531; AAD20387.2; JOINED.
 DR EMBL; AF177532; AAD20387.2; JOINED.
 DR EMBL; AF177533; AAC02527.2; JOINED.
 DR EMBL; AF043196; AAC02527.2; JOINED.
 DR EMBL; AF043197; AAC02527.2; JOINED.
 DR EMBL; AF177518; AAC02527.2; JOINED.
 DR EMBL; AF177519; AAC02527.2; JOINED.
 DR EMBL; AF177520; AAC02527.2; JOINED.
 DR EMBL; AF177521; AAC02527.2; JOINED.
 DR EMBL; AF177522; AAC02527.2; JOINED.
 DR EMBL; AF177523; AAC02527.2; JOINED.
 DR EMBL; AF177524; AAC02527.2; JOINED.
 DR EMBL; AF177525; AAC02527.2; JOINED.
 DR EMBL; AF177526; AAC02527.2; JOINED.
 DR EMBL; AF177527; AAC02527.2; JOINED.
 DR EMBL; AF177528; AAC02527.2; JOINED.
 DR EMBL; AF177529; AAC02527.2; JOINED.
 DR EMBL; AF177530; AAC02527.2; JOINED.
 DR EMBL; AF177531; AAC02527.2; JOINED.

| | | | | | | |
|----|---|---|----------|---|--|---------------------------|
| DR | EMBL; AF177532; AAC02527.2; JOINED. | FT | CONFLICT | 996 | 996 | P -> S (IN REF. 1). |
| DR | EMBL; AF177533; AAD56218.2; JOINED. | FT | CONFLICT | 1092 | 1095 | ARLQ -> GPGL (IN REF. 1). |
| DR | EMBL; AF043195; AAD56218.2; JOINED. | FT | CONFLICT | 1136 | 1136 | S -> N (IN REF. 5). |
| DR | EMBL; AF043196; AAD56218.2; JOINED. | FT | CONFLICT | 1155 | 1158 | GSYG -> RSFC (IN REF. 5). |
| DR | EMBL; AF177518; AAD56218.2; JOINED. | FT | CONFLICT | 1165 | 1167 | EYR -> IRS (IN REF. 5). |
| DR | EMBL; AF177519; AAD56218.2; JOINED. | SQ | SEQUENCE | 1190 | AA; 133971 MW; BE2BE6F181467058 CRC64; | |
| DR | EMBL; AF177520; AAD56218.2; JOINED. | Query Match | | | | |
| DR | EMBL; AF177521; AAD56218.2; JOINED. | Best Local Similarity | | | | |
| DR | EMBL; AF177522; AAD56218.2; JOINED. | Matches 213; Conservative 144; Mismatches 394; Indels 239; Gaps 50; | | | | |
| DR | EMBL; AF177523; AAD56218.2; JOINED. | QY | 55 | EEVLHSPRLTNSAMRAGHLDDLKTRKNGKATAFLESKFHNP--DVYTLVTGLQPDVDF | 112 | |
| DR | EMBL; AF177524; AAD56218.2; JOINED. | Db | 94 | EDVLSH-FAVQQLRKSGKVAIVVPRKRVQVAALQA----SPPLQDDRAFFVMEFQ | 148 | |
| DR | EMBL; AF177525; AAD56218.2; JOINED. | QY | 113 | SNF-SGLMETSKLTECLAGALSLOELAQEKGQKEVLLRRCCOQLQELHGLAETRAEGH | 171 | |
| DR | EMBL; AF177526; AAD56218.2; JOINED. | Db | 149 | RSFRSGYSERSRLNS-HGGRSRSWED--SPERGPH---ERASRRDLSRDSRGRSLE | 202 | |
| DR | EMBL; AF043197; AAD56219.2; JOINED. | QY | 172 | Q-LEADHSRMKREVSAAHFHEVLRLKDEMLSLSL-HYSNALQEKELAAASRCRSLEQEL | 229 | |
| DR | EMBL; AF177518; AAD56219.2; JOINED. | Db | 203 | RGLDQDHARTD-----RSRGRSLERGLDHDGFSRDRDRDRSRGRSID----- | 246 | |
| DR | EMBL; AF177519; AAD56219.2; JOINED. | QY | 230 | KQELQRANMVSCLELELQFOSLRTASDQSGDEELNRLKEENKLSLTSFLAEKILEQ | 289 | |
| DR | EMBL; AF177520; AAD56219.2; JOINED. | Db | 247 | -QDYERAY-----HRAYPDYER----- | 265 | |
| DR | EMBL; AF177521; AAD56219.2; JOINED. | QY | 290 | SILDEARGSRQELVERTHSLRERAVAAERQYWEKEQTLLQFQSKMACOLYREKVA | 349 | |
| DR | EMBL; AF177522; AAD56219.2; JOINED. | Db | 266 | SPEYRRGAR-----HDARSRG-PRSRREHPHSRSPSPPEPRGPGICVLMLKSRAN- | 316 | |
| DR | EMBL; AF177523; AAD56219.2; JOINED. | QY | 350 | LOAQVCELOKQDAYSDASDAQ---REISQS-LVEKDSLRQ---VFELTDQVCE--LR | 400 | |
| DR | EMBL; AF177524; AAD56219.2; JOINED. | Db | 317 | -----EYGLRLGSOIFVKEMTRTGLATKQGNLHEGDIILKINGVTENMSL | 363 | |
| DR | EMBL; AF083892; AAC33121.1; JOINED. | QY | 401 | TQLRQLOAEPGPGVLKQEAETRECPREKQPLVRMHAICPRDDSDCL--VSSTES | 453 | |
| DR | EMBL; AF083893; AAC33122.1; JOINED. | Db | 364 | TDARKLIEKSRGKL-QLVVLRD---SQQTLLINIPSL---NDSDEIETISESNRFS | 415 | |
| DR | EMBL; U84581; AAB41794.1; JOINED. | QY | 454 | -----QLLSDLSATSSRELVDSPRSSPAPPSSOOLYKRVAEDEGPEPWFSSCLEI | 508 | |
| DR | HSSP; P31016; IBE9. | Db | 416 | PEERHOYSDDYHSSEKLERPSSREDTPSRLS-----RMGATPTFKS----- | 461 | |
| DR | InterPro; IPR000619; Guanylate_kin. | QY | 509 | DPGALPGAKAGDPHLDYELDDTADLPQLESSLPQSPGRLDVSESGVIMRRRRPARRL-- | 566 | |
| DR | InterPro; IPR001478; PDZ. | Db | 462 | -TGDIAGTVVPTNKPRYQEEPPAP-----QPKAAPR-----TFLRPSPEDEAI | 507 | |
| DR | Pfam; PF00625; Guanylate_kin; 1. | QY | 567 | SQVTMLAF-QGDALLEQISVIGGNLTGIFHRTVTPGSAADQMALRPGTQIVMVDYASEP | 625 | |
| DR | Pfam; PF00595; PDZ; 3. | Db | 508 | PNTKMYRFRKGDV--GLRLAGGNDVGFVAGIQEGTSABEQGLQEGDQILKVTQDFRG | 565 | |
| DR | SMART; SM00072; Gukc; 1. | QY | 626 | LFKAVLEDTTLEAVGLLRVDG-FCCLSVKVNTDGYKRLQLDLEAKVATSGDSFYIRVN | 684 | |
| DR | SMART; SM00228; PDZ; 3. | Db | 566 | L-----VREDAVLY-----LLEIPKGMVTLAQSRADVTVDIL-----ACRGDS | 612 | |
| DR | SMART; SM00326; SH3; 1. | QY | 685 | LAMEGRAKGELQVHCNEVLHVTDMFQG-CGCWHAHVRNVTYMKDTAAHGTIPNYSRAQ | 743 | |
| DR | PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG. | Db | 613 | FECEKETPQSLAFTRGVFRVVDLYDGKLGNLAVRIGNELEK-----GLIPNKSRAEQ | 667 | |
| DR | PROSITE; PS00552; GUANYLATE_KINASE_2; 1. | QY | 744 | QLIALQDWTQOCTVTTRKPSGGGPKLVRIVSMDKAKASPLRSLFRDGLQDPSRMESST | 803 | |
| DR | PROSITE; PS50106; PDZ; 3. | Db | 668 | --MASVQN-----AQRDNAGDRADFWM-----RGORSVGKKNLRKSRDLTA | 714 | |
| DR | Tight Junction; SH3 domain; Repeat; Membrane; Alternative splicing; Alternative initiation. | QY | 804 | CFWAESCLTLVPTLVWPHRPARPRVLLVPRVAGKLTSLSEKLCILQGLKCLAEVLSQEE | 863 | |
| DR | 1 | Db | 715 | KFFA-----YERVLLREAGFKRPVLF-GPIADTAMEKLA-----NE | 750 | |
| DR | 2 | QY | 864 | YEAWSRGDIIOGEYSGGRCWVTR-HAVESLMKNTNHALDVLDSVCTLHRMDIPPIV | 922 | |
| DR | 3 | Db | 751 | LPDWFQTAKEPKDAGSEKSTGVVRLNTVQIIQDQKHALDVTTPKAVDLLNTYQWPIV | 810 | |
| DR | 4 | QY | 923 | IHVSVNEKMAKLLKGLQRL-----GTSEEQLLEAARQEGDLDRAPCLY-----SSLA | 971 | |

```

Db 811 IFFNPDSRGQVKMK--QRLNPTSNKSRKLFDOANKL-----KKTCAHLFTATINLNSA 863
QY 972 PDGWSLDGLLSCVRCVRAIADEQKQVWTEQ 1001
Db 864 NDSW-----FGSLKDTIQHQGEAVWSE 887

RESULT 13
PLEL_HUMAN
ID PLEL_HUMAN STANDARD; PRT; 4684 AA.
AC Q15149; Q16640; Q15148;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Plectin 1 (PLTN) (PCN) (Hemidesmosomal protein 1) (HD1).
GN PLECI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Placenta;
RX MEDLINE=96210632; PubMed=8633055;
RA Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;
RT "Human plectin: organization of the gene, sequence analysis, and
RT chromosome localization (8q24).";
RL Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND DISEASE.
RX MEDLINE=96312447; PubMed=8698233;
RA McLean W.H.I., Pulkkinen L., Smith F.J.D., Rugg E.L., Lane E.B.,
RA Bullrich F., Burgeson R.E., Amano S., Hudson D.L., Owaribe K.,
RA McGrath J.A., McMillan J.R., Eady R.A.J., Leigh I.M., Christiano A.M.,
RA Uitto J.;
RT "Loss of plectin causes epidermolysis bullosa with muscular dystrophy:
RT cDNA cloning and genomic organization.";
RL Genes Dev. 10:1724-1735(1996).
RN [3]
RP VARIANT MD-EBS 1003-GLN--ALA-1005 DEL.
RX MEDLINE=97049959; PubMed=8894687;
RA Pulkkinen L., Smith F.J.D., Shimizu H., Murata S., Yaoita H.,
RA Hachisuka H., Nishikawa T., McLean W.H.I., Uitto J.;
RT "Homozygous deletion mutations in the plectin gene (PLEC1) in patients
RT with epidermolysis bullosa simplex associated with late-onset
RT muscular dystrophy.";
RL Hum. Mol. Genet. 5:1539-1546(1996).
RN [4]
RP VARIANT MD-EBS LEU-429 INS.
RX MEDLINE=21090821; PubMed=11159198;
RA Bauer J.W., Rouan F., Kofler B., Reznicek G.A., Kornacker I.,
RA Muss W., Hametner R., Klaussegger A., Huber A., Pohla-Gubo G.,
RA Wiche G., Uitto J., Hentner H.;
RT "A compound heterozygous one amino-acid insertion/nonsense mutation in
RT the plectin gene causes epidermolysis bullosa simplex with plectin
RT deficiency.";
RL Am. J. Pathol. 158:617-625(2001).
CC -1- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND
CC MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSES OR
CC HEMIDESMOSOMES. COULD ALSO BIND MUSCLE PROTEINS SUCH AS ACTIN TO
CC MEMBRANE COMPLEXES IN MUSCLE. MAY BE INVOLVED NOT ONLY IN THE
CC CROSSLINKING AND STABILIZATION OF CYTOSKELETAL INTERMEDIATE
CC FILAMENTS NETWORK, BUT ALSO IN THE REGULATION OF THEIR DYNAMICS.
CC -1- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST LEVELS IN
CC MUSCLE, HEART, PLACENTA AND SPINAL CORD.
CC -1- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN. THE C-TERMINUS WITH
CC VIMENTIN, DESMIN, GFAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-
CC AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.
CC -1- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM

```

```

CC INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).
CC DISEASE: DEFECTS IN PLECI ARE THE CAUSE OF EPIDERMOLYSIS BULLOSA
CC WITH MUSCULAR DYSTROPHY (MD-EBS OR EB-MD), AN AUTOSOMAL RECESSIVE
CC DISORDER CHARACTERIZED BY EPIDERMAL BLISTER FORMATION AT THE LEVEL
CC OF THE HEMIDESMOSOME AND ASSOCIATED LATE-ONSET MUSCULAR DYSTROPHY.
CC -1- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC -1- SIMILARITY: CONTAINS 33 PLECTIN REPEATS.
CC -1- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.
CC -1- SIMILARITY: BELONGS TO THE PLEKIN OR CYTOLINKER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; Z54367; CAA91196.1; -
CC EMBL; U53204; AAB05427.1; -
CC EMBL; U63610; AAB05428.1; -
CC EMBL; U63609; AAB05428.1; JOINED.
CC EMBL; X97053; CAA65765.1; -
CC HSSP; Q01082; 1BKR.
CC MIM; 601282; -
CC MIM; 226670; -
CC InterPro; IPR001589; Actinin_act_bind.
CC InterPro; IPR001715; Calponin_hom.
CC InterPro; IPR001101; Plectin_repeat.
CC InterPro; IPR002017; Spectrin.
CC Pfam; PF00307; CH; 2.
CC Pfam; PF00681; Plectin_repeat; 19.
CC SMART; SM00033; CH; 2.
CC SMART; SM00250; PLEC; 33.
CC SMART; SM00150; SPEC; 5.
CC PROSITE; PS00019; ACTININ_1; FALSE_NEG.
CC PROSITE; PS00020; ACTININ_2; FALSE_NEG.
CC PROSITE; PS50021; CH; 2.
CC Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;
CC Phosphorylation; Alternative splicing; Epidermolysis bullosa;
CC Disease mutation.
CC DOMAIN 1 1470
CC DOMAIN 1471 2755
CC DOMAIN 2756 4684
CC DOMAIN 175 400
CC DOMAIN 179 282
CC DOMAIN 295 397
CC DOMAIN 645 710
CC REPEAT 740 824
CC REPEAT 837 930
CC REPEAT 1315 1415
CC DOMAIN 1469 2756
CC REPEAT 2826 2863
CC REPEAT 2864 2901
CC REPEAT 2902 2939
CC REPEAT 2940 2977
CC REPEAT 2981 3015
CC REPEAT 3116 3153
CC REPEAT 3154 3191
CC REPEAT 3192 3229
CC REPEAT 3230 3267
CC REPEAT 3268 3305
CC REPEAT 3306 3343
CC REPEAT 3485 3522
CC REPEAT 3523 3560
CC REPEAT 3561 3598
CC REPEAT 3599 3636
CC REPEAT 3640 3674
CC REPEAT 3820 3857
CC REPEAT 3858 3895
CC REPEAT 3896 3933
CC REPEAT 3934 3971

```


RX MEDLINE-92003925; PubMed-1912569;
 RA Toothaker L.E., Gonzalez D.A., Tung N., Lemons R.S., le Beau M.M.,
 RA Arnaout M.A., Clayton L.K., Tenen D.G.;
 RT "Cellular myosin heavy chain in human leukocytes: isolation of 5'
 RT cDNA clones, characterization of the protein, chromosomal
 RT localization, and upregulation during myeloid differentiation.";
 RL Blood 78:1826-1833(1991).
 RN [3]
 RP SEQUENCE OF 1-715 FROM N.A.
 RX MEDLINE-91316803; PubMed-1860190;
 RA Sinons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,
 RA Gulan D., Adelstein R.S., Weir L.;
 RT "Human nonmuscle myosin heavy chains are encoded by two genes located
 RT on different chromosomes.";
 RL Circ. Res. 69:530-539(1991).
 RN [4]
 RP SEQUENCE OF 714-1960 FROM N.A.
 RX MEDLINE-90138958; PubMed-1967836;
 RA Saez C.G., Myers J.C., Shows T.B., Leinwand L.A.;
 RT "Human nonmuscle myosin heavy chain mRNA: generation of diversity
 RT through alternative polyadenylation.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1164-1168(1990).
 RN [5]
 RP VARIANT DFNA17 HIS-705.
 RX MEDLINE-20489856; PubMed-11023810;
 RA Lalwani A.K., Goldstein J.A., Kelley M.J., Luxford W., Castelein C.M.,
 RA Mhatre A.N.;
 RT "Human nonsyndromic hereditary deafness DFNA17 is due to a mutation in
 RT nonmuscle myosin MYH9.";
 RL Am. J. Hum. Genet. 67:1121-1128(2000).
 RN [6]
 RP VARIANTS MHA/FTNS/SBS K-93; C-702; C-1165; H-1424 AND K-1841.
 RX MEDLINE-20428192; PubMed-10973259;
 RA Seri M., Cusano M., Gangarossa S., Caridi G., Bordo D., Lo Nigro C.,
 RA Giggeri G.M., Ravazzolo R., Savino M., Del Vecchio M., d'Apolito M.,
 RA Iolascon A., Zeltante L.B., Savio A., Balduini C.L., Norris P.,
 RA Magrini U., Belletti S., Heath K.E., Babcock M., Glucksman M.J.,
 RA Aliprandis E., Bizzaro N., Deanick R.J., Martignetti J.A.;
 RT "Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and
 RT Sebastian syndromes.";
 RL Nat. Genet. 26:103-105(2000).
 RN [7]
 RP VARIANTS MHA ILE-1155 AND LYS-1841.
 RX MEDLINE-20428193; PubMed-10973260;
 RA Kelley M.J., Jawien W., Ortel T.L., Korczak J.F.;
 RT "Mutation of MYH9, encoding non-muscle myosin heavy chain A, in
 RT May-Hegglin anomaly.";
 RL Nat. Genet. 26:106-108(2000).
 CC -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
 CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
 CC CAPPING.
 CC -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
 CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
 CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF MAY-HEGGLIN ANOMALY
 CC (MHA), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
 CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.
 CC -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF FECHTNER SYNDROME
 CC (FTNS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
 CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.
 CC WITH ADDITIONAL ALPORT-LIKE CLINICAL FEATURES OF SENSORINEURAL
 CC DEAFNESS, CATARACTS AND NEPHRITIS.
 CC -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF SEBASTIAN SYNDROME
 CC (SBS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
 CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.
 CC -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF AN AUTOSOMAL DOMINANT
 CC FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNA17) WHICH IS
 CC CHARACTERIZED BY PROGRESSIVE HEARING IMPAIRMENT AND
 CC COCHLEOSACULAR DEGENERATION.
 CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.

CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; 282215; CAB05105.1; -;
 DR EMBL; M81105; AAA59888.1; -;
 DR EMBL; M69180; AAA61765.1; -;
 DR EMBL; M31013; AAA36349.1; -;
 DR HSSP; P08799; 1LVK.
 DR MIM; 160775; -;
 DR MIM; 153640; -;
 DR MIM; 155100; -;
 DR MIM; 603622; -;
 DR MIM; 605249; -;
 DR InterPro; IPR000048; IQ.
 DR InterPro; IPR004009; Myosin_N.
 DR InterPro; IPR002928; Myosin_tail.
 DR InterPro; IPR002017; Spectrin.
 DR Pfam; PF00612; IQ; 1.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF02736; Myosin_N; 1.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS00096; IQ; 1.
 KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
 KW Coiled coil; Alkylation; Multigene family; Disease mutation;
 KW Deafness.
 FT DOMAIN 1 778 MYOSIN HEAD-LIKE.
 FT DOMAIN 779 808 IQ.
 FT DOMAIN 837 1926 COILED COIL (POTENTIAL).
 FT NP_BIND 174 181 ATP (POTENTIAL).
 FT DOMAIN 654 676 ACTIN-BINDING.
 FT MOD_RES 704 704 ALKYLATION (SH-1) (POTENTIAL).
 FT VARIANT 93 93 ALKYLATION (SH-2) (POTENTIAL).
 FT VARIANT 702 702 N -> K (IN MHA).
 FT VARIANT 702 702 /FTID=VAR_010791.
 FT VARIANT 705 705 R -> C (IN FTNS).
 FT VARIANT 705 705 /FTID=VAR_010792.
 FT VARIANT 1155 1155 R -> H (IN DFNA17).
 FT VARIANT 1155 1155 /FTID=VAR_010793.
 FT VARIANT 1165 1165 T -> I (IN MHA).
 FT VARIANT 1424 1424 /FTID=VAR_010794.
 FT VARIANT 1424 1424 R -> C (IN SBS).
 FT VARIANT 1841 1841 /FTID=VAR_010795.
 FT CONFLICT 53 55 D -> H (IN FTNS).
 FT CONFLICT 660 660 E -> K (IN MHA).
 FT CONFLICT 869 869 /FTID=VAR_010797.
 FT CONFLICT 931 931 EAI -> RGH (IN REF. 3).
 FT CONFLICT 1240 1241 T -> S (IN REF. 3).
 FT CONFLICT 1350 1350 C -> Y (IN REF. 4).
 FT CONFLICT 1764 1764 KG -> GR (IN REF. 4).
 FT CONFLICT 1771 1771 E -> EE (IN REF. 2).
 FT CONFLICT 1771 1771 T -> A (IN REF. 2).
 FT CONFLICT 1771 1771 S -> G (IN REF. 2).
 SQ SEQUENCE 1960 AA; 226531 MW; 588F84BB8C10656F CRC64;

Query Match 4.7%; Score 240.5; DB 1; Length 1960;
 Best Local Similarity 21.1%; Pred. No. 0.00014;
 Matches 238; Conservative 176; Mismatches 359; Indels 353; Gaps 59;

```
QY 14 LDETLWEMMESHRH---RIVRCIPS-----RLTPYLQAKVLCOLDDEEVHLSPR-- 62
Db 649 LYEQALAKMATLNTNPNFVRCIIPNHEKKAGKLDPHL-----VDLDRONGVLEGIRIC 704
QY 63 -----LITNSMRAGHLLDLKTRGKNGAIATFLESKLFHNPVYTL- 102
Db 705 RQGFNPNRVVFQFRQRYEILTPNSIPKGF-----DGKQACVLMIKALEL-DSNLYRIG 757
QY 103 -----VTGLQPDVDFSNFSGLMETSKTTECLAG-----AIGSLQEEFN 140
Db 758 QSKVFFRAGVLAHLEERDL-----KITDVIIGFOACCGRGYLARKAFARQQOLT 807
QY 141 QEGQKEVILLRCCQQLQEHGLAE-----TRAEGHLQLEADHSRMKRVSAHFHEVLR 193
Db 808 AMK-----VLRQNC---AAYLKLNNQWMLFTKVKPLQV-----SQEEMMAKEEELVK 856
QY 194 LKDEMLSLSHYSNALQEKELAAASRCRQEEELYLLKQELQORANMVSSCELEQEQLSR- 252
Db 857 VREKQLAA---ENRLTME-----TLOSQMAEKLQLOE-QLQAEYELCAEAELRA 904
QY 253 --TASQESGDEBL-----NRLKEENEKLSLTFSLAEKDILEQSLDEARGSQELVERI 305
Db 905 RLTAKKQEL--EEICHDLEARVEEERCOHLQ---AERKKMQQNIQELEEQUEEESAR 959
QY 306 HSLRERAAAREOREQYWEKEOTLLQFOKSKMACOLYREKYNALQAOYCE----LQKER 361
Db 960 QKQLEKVTYEAKKKL--EEQIILLEDQ---CKLAKK-KLEDRIAFTNTLTEE 1012
QY 362 DQAYS-----ARDSAQR---EISQSLVERKDS--LRQVPFELDQVC 397
Db 1013 EKSKSLAKLKNKHEAMITDLEERLRREEQRQLEKTRRKLEGSDTDLSDQIAELQAQIA 1072
QY 398 ELRTQLRQLOQAEPPGVYLQEAFTREPCPREKQRLVRMHAICPRDDSDCSLVSSSTESQLLS 457
Db 1073 ELKMQLAKEEELQOAL---ARVEEAAQKNMALKKIREL-----ESQ-IS 1114
QY 458 DLSATSSRELVSFRSSPAPPQQSLYKRVADFGCEPWSFSSCLEIPEGDPGALPGAK 517
Db 1115 EL-----QEDLESERASRNKAQKQ-----RDLGEE-----LEA----- 1143
QY 518 AGDPHLDYELLDTADLPQLESSLPQVSPGRLDVSESGV-----LMRRRPARRI 565
Db 1144 -----LKTELEDTLDTAAQQELRSKREQEVNLLKTLTEEAKEAQTQEMRQKHSQ-- 1196
QY 566 LSQVTMLAFQGDALLRQISVIGNL-----TGIFTHRV-----TPGSAADOMAL 609
Db 1197 --AVEELAQ-----LEQTRKRVANLEKAKOTLENERGELANEVYLLQKGDSHKKRV 1250
QY 610 RPTQIIVMYDYEASEPLKAVLEDDT-----LEEAVGLLRVRDGFCCLSVKVNTD--GYK 662
Db 1251 EAQLOELQVKFNEGERVTRTELADKVTKLQVELDNVTGLLSQSDS---KSKLTKDFSAL 1307
QY 663 RLQDLE-----AKVATSGDSFYIRVNLAMEGRAKGBELQ-----VHCN 700
Db 1308 SOLQDTQELQEBENRQKLSLTKLQKQVEDEKNSF--REQLEEEEAKNHLEKQIATLHA- 1364
QY 701 EVLHVTDTMFQCGCWH--AHRVNSYTMKD-----TAAHGTI-PNYSRAQQOLIA 747
Db 1365 QVADMKKKMEDSVGCLETAEVYKLRQKLEGLSQRHEEKVAAQYDKLEKTKRLQOELDD 1424
QY 748 LIQDMTQ-----CTVTRKPSGGPQKLVIRVSMDAKASPLRLSDFRGOLDPSRMEGSST 803
Db 1425 LLVDLDHQRSACNLEKK-----QKKFDQLAAEKTISAKYAEERDRAEAAREKETKA- 1478
QY 804 CFWAESCLTLVPYTLVWPHRPARPRVLLVPRANGKILSEKLCILQGFKKCLAEYLSQEE 863
Db 1479 -----LSLARALEAMEQKAEEL-----ERLNKQPRTEM 1506
QY 864 YEAWSORGDI---IOBGEVSGRCWVTRHAVESLMEKNTHALLDVOLDVSVCTLHRMDIFP 920
Db 1507 EDLMSSKDDVGKSVHELEKS-----KRALEQQVEEMKTQLEELEDELOATEDAK---- 1555
QY 921 IVIHVSVN-EKMAKKLKGLO-RLGTSEE---QLLEAARQEEGDLD 961
```

Db 1556 --LRLEVNLOAKMAQOAFERDLOGRDEQSEKKKQOLVRQVREMEAELE 1599

Search completed: June 13, 2002, 09:26:01
Job time: 184 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2002, 09:22:22 ; Search time 26.06 Seconds
(without alignments)
3701.984 Million cell updates/sec

Title: US-09-767-215-2

Perfect score: 5149

Sequence: 1 MGELCRDSALTALDEETLW.....VROAIADQKVKVWTEQSPR 1004

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 289 | 5.6 | 1745 | 2 A46431 | tight junction-ass |
| 2 | 282 | 5.5 | 1367 | 2 T13703 | tama protein - fru |
| 3 | 270.5 | 5.3 | 1736 | 2 A47747 | tight junction pro |
| 4 | 262 | 5.1 | 2101 | 2 A42184 | nuclear mitotic ap |
| 5 | 249 | 4.8 | 4574 | 2 G02520 | plectin - human |
| 6 | 248.5 | 4.8 | 4684 | 2 A59404 | plectin [imported] |
| 7 | 248.5 | 4.8 | 2442 | 2 T08621 | centrosome associa |
| 8 | 246.5 | 4.8 | 1116 | 2 I54378 | gene x104 protein |
| 9 | 245 | 4.8 | 746 | 2 T47237 | myosin II heavy ch |
| 10 | 243.5 | 4.7 | 2168 | 2 T30171 | ninein - mouse |
| 11 | 242.5 | 4.7 | 3225 | 2 I52300 | giantin - human |
| 12 | 242.5 | 4.7 | 3259 | 1 A56539 | giantin - human |
| 13 | 242 | 4.7 | 1163 | 2 JE0366 | tight junction pro |
| 14 | 236.5 | 4.6 | 775 | 2 I46236 | tight junction pro |
| 15 | 235 | 4.6 | 1959 | 1 A33977 | myosin heavy chain |
| 16 | 235 | 4.6 | 1961 | 1 A61231 | myosin heavy chain |
| 17 | 234.5 | 4.6 | 4687 | 1 A39638 | plectin - rat |
| 18 | 230 | 4.5 | 1281 | 2 T00346 | myosin heavy chain |
| 19 | 226 | 4.4 | 1972 | 1 A41604 | myosin heavy chain |
| 20 | 225.5 | 4.4 | 2094 | 2 S33124 | tpr protein - huma |
| 21 | 225 | 4.4 | 1017 | 2 PC4035 | cell-cycle-depende |
| 22 | 223 | 4.3 | 1286 | 2 T16507 | hypothetical prote |
| 23 | 222.5 | 4.3 | 1133 | 2 T22976 | hypothetical prote |
| 24 | 222 | 4.3 | 876 | 2 A23767 | myosin heavy chain |
| 25 | 222 | 4.3 | 2649 | 2 A40937 | bullous pemphigoid |
| 26 | 219 | 4.3 | 1940 | 2 A59287 | myosin heavy chain |
| 27 | 218.5 | 4.2 | 1388 | 2 S74245 | serine/threonine-s |
| 28 | 218.5 | 4.2 | 1955 | 2 T30934 | myosin-like protei |
| 29 | 218.5 | 4.2 | 3187 | 2 JC5837 | 364K Golgi complex |

| | | | | | |
|----|-------|-----|------|----------|--------------------|
| 30 | 217.5 | 4.2 | 1976 | 2 A59252 | myosin heavy chain |
| 31 | 217 | 4.2 | 1938 | 2 JC5421 | smooth muscle myos |
| 32 | 217 | 4.2 | 1972 | 2 JC5420 | smooth muscle myos |
| 33 | 216.5 | 4.2 | 1388 | 2 S70633 | serine/threonine-s |
| 34 | 215.5 | 4.2 | 1034 | 2 T22166 | hypothetical prote |
| 35 | 215.5 | 4.2 | 1091 | 2 T34107 | hypothetical prote |
| 36 | 215 | 4.2 | 1837 | 2 T41023 | probable nuclear p |
| 37 | 215 | 4.2 | 1974 | 2 T30010 | hypothetical prote |
| 38 | 214 | 4.2 | 1407 | 1 S28589 | trichohyalin - rab |
| 39 | 213.5 | 4.1 | 1927 | 1 A59236 | embryonic muscle m |
| 40 | 213 | 4.1 | 2253 | 2 T30336 | nuclear/mitotic ap |
| 41 | 212.5 | 4.1 | 1940 | 1 A24922 | myosin heavy chain |
| 42 | 212 | 4.1 | 1957 | 2 A59294 | skeletal myosin - |
| 43 | 212 | 4.1 | 1957 | 2 T38077 | hypothetical colle |
| 44 | 212 | 4.1 | 2954 | 2 T14156 | kinesin-related pr |
| 45 | 211.5 | 4.1 | 1218 | 2 T14265 | golgin-245 - mouse |

ALIGNMENTS

RESULT 1

A46431

tight junction-associated protein ZO-1 - mouse

C:Species: Mus musculus (house mouse)

C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999

C:Accession: A46431

R:Itch, M.; Nagafuchi, A.; Yonemura, S.; Kitani-Yasuda, T.; Tsukita, S.; Tsukita, S.

J. Cell Biol. 121, 491-502, 1993
A:Title: The 220-kD protein colocalizing with cadherins in non-epithelial cells is id
py.

A:Reference number: A46431; MUID:93252986

A:Accession: A46431

A:Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-1745 <I/O>

A:Cross-references: GB:D14340; NID:g303709; PIDN:BAA03274.1; PID:d1003784; PID:g30371

A:Experimental source: F9 cells

A:Note: sequence extracted from NCBI backbone (NCBIN:131200, NCBTP:131201)

C:Superfamily: guanylate kinase homology; GLGF domain homology

F:27-106/Domain: GLGF domain homology <GLG1>

F:428-498/Domain: GLGF domain homology <GLG3>

F:645-794/Domain: guanylate kinase homology <GKI>

Query Match 5.6% Score 289; DB 2; Length 1745;
Best Local Similarity 22.3%; Pred. No. 2.5e-07;
Matches 191; Conservative 136; Mismatches 319; Indels 212; Gaps 43;

Qy 231 QELQRANVVSCELELOFQSLRTASDQSGDEELNRLKEENEKRLSLTFFSLAEKDILRQS 290

Db 70 QENDRVAMVNGVSMNVYEHAFVQOLRKSQKNA--KITIRKKKKVQIPVSHPDPEVSDN 127

Qy 291 LDEARGSRQELVERIHSLRE--RAVAERQROYWEEKETLLQFOKSKMACOLYREKVA 349

Db 128 EDDSYD-----EEVHDPRAGRGALANRRSEKSWARDRSA---SRERSLSRSPRRSVAS 178

Qy 350 LQ-----AQVCELOKEROQVARSQAQ---REISO-SLVEKDSLRRQ---VFELTDQVCE 398

Db 179 SQAKPTKVTLVKRRKNEEYGLRPASHIFVKEISQDSLAAARDGDTQEGVVKLTNGVTYE 238

Qy 399 --LRTQLRQLQAEPGPGVLKQEARTEPCPREKQRLVR-----MHAICPRDDSDCSLVS 449

Db 239 NMSLTDAKTLIERSGKGLKVVQR-----DERATLLNVPDLSDSIHSANASERODISEIQ 293

Qy 450 STESQLLSLDSATS-----SRELVSFRSSSPAPPSQOSLYKRVADFEE 495

Db 294 S-----LASHSGRGRHPPRRRSQSRSPQDRSEPDSHTSQSPQPSNGSLRSR-EEERMSK 348

Qy 496 PWFSSSCLE-----IPEGDPGALPGAKAGDPHLDYELDDTA 531

Db 349 PGAISTPVKKVHDDHPKKAEEVTVKEKNEKQPTLPPEPKP-----VYAAQVGQPDVL----- 399

| | | | | |
|----|-----|---|---|-----|
| Qy | 542 | PVSPGR | ---LDV-----SESVLM-----RRPARRILSVQTMALAFQGDALL | 580 |
| | | : | : | |
| Db | 358 | PVTRGRSAAIDPPRPPPPRSGGAQEDFYSSRQLYEERQSAEPRFISFQREGSV | 417 | |
| | | : | : | |
| Qy | 581 | EQISVIGNLTGFIHVRTPGSAAQOMALPGGTQIVMVDYEASEPLFKAVLEDTTLEEAV | 640 | |
| | | : | : | |
| Db | 418 | -GIRLTGCGNAGIPVAVQPGSPASLQGLMPGDKILKN-----DMDMCGVTREEAV | 468 | |
| | | : | : | |
| Qy | 641 | GLLRRVDGFCCLSVKVNVDGYKRLQLDLEAKVATS--GGSFYIRVNLMMEGRAKGELOVH | 698 | |
| | | : | : | |
| Db | 469 | LFL-----LSLQDRIDLIVQYCKEYDEVVYTNORGDSFHILKTHFCDNPSKSGEMAFK | 520 | |
| | | : | : | |
| Qy | 699 | CNEVLHVTDTMFOG-CGCHWAHRVNSYTMKDTAAHGTIPNYSRAAQOOLIALIQDMTQOCT | 757 | |
| | | : | : | |
| Db | 521 | AGDVERVIDTLHVGVSQVWLKIGRCHQE--MQGVIPNKSRAEEL-----ATAQFN | 571 | |
| | | : | : | |
| Qy | 758 | VTRKPSGGPOKLVRIVSMOKAKASPLRLSFDRGQLDPSR-----WEGSSTCFWAEESCL | 811 | |
| | | : | : | |
| Db | 572 | ATKK-----EMANESRGHFFRRRSTHRRSKLSRENMDVDVVFSDSIS | 615 | |
| | | : | : | |
| Qy | 812 | TLVPTYLWPHRPARPRPVLVVPRAVGKILSEKLCLOGEKKCLAEVLSQEEYEAWSQRG | 871 | |
| | | : | : | |
| Db | 616 | KFPAYERVLRHGCFVRPVLF-GPYSDLABERLA--KDFPKFSPLOQDDKSA----- | 667 | |
| | | : | : | |
| Qy | 872 | DIQEGEVSGRCRWTRHA--VESLMEKNTHALLDVLQDSVCTTLHRMDIPPIVTHVSNEK | 930 | |
| | | : | : | |
| Db | 668 | -----ATSCKCRIVLSNIRDMYDRGKHALLDITPNVADRLNVAQFYPVWIFLKTDSK | 720 | |
| | | : | : | |
| Qy | 931 | -MAKKLKKGLQRLG-TSEEGOLLEAAREEGDRLAPCLYSLLAPDGW-----SDL | 978 | |
| | | : | : | |
| Db | 721 | HVHKQLRHGLPKAAHKSSKLLLECQCK---LERV-----WSHIFSTQIALSDE | 765 | |
| | | : | : | |
| Qy | 979 | DGLLSCVRQAIADEQKKVWVTEQS | 1002 | |
| | | : | : | |
| Db | 766 | ESWYRKLRDSDIDLOQSGAVWMS | 789 | |
| | | : | : | |

```

RESULT      3
A47747
    tight junction protein ZO-1 - human
C:Species: Homo sapiens (man)
C:Date: 02-Jun-1994 #sequence_revision 02-Jun-1994 #text_change 05-Nov-1999
C:Accession: A47747
R:Willott, E.; Balda, M.S.; Fanning, A.S.; Jameson, B.; Van Itallie, C.; A
Proc. Natl. Acad. Sci. U.S.A. 90, 7834-7838, 1993
A:Title: The tight junction protein ZO-1 is homologous to the Drosophila d
A:Reference number: A47747; MUID:93361541
A:Accession: A47747
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1736 <MIL>
A:Cross-references: GB:L14837; NID:g292937; PIDN:AAA02891.1; PID:g292938
C:Superfamily: guanylate kinase homology; GLGF domain homology
C:Keywords: alternative splicing; membrane protein; phosphoprotein
F:15-94/Domain: GLGF domain homology <GLG1>
F:181-248/Domain: GLGF domain homology <GLG2>
F:416-486/Domain: GLGF domain homology <GLG3>
F:633-782/Domain: guanylate kinase homology <GKI>

```

F:633-782/Domains: guanylate kinase homology <GKI>

| | Query Match | 5.3% | Score 270.5; | DB 2; | Length 1736; |
|----|--|-------------------|--------------------|-------------|--------------|
| | Best Local Similarity | 21.4%; | Pred. No. 2.4e-06; | | |
| | Matches 174; | Conservative 129; | Mismatches 278; | Indels 233; | Gaps |
| Qy | 247 QEOSLRTASDQSGDEE-----LNRLKKEENXLRSLTFSIAEKDTLEOSLSDEARG | 296 | | | |
| | : : : : | : | : | : : : | |
| Db | 151 RERSLSPRSDRRSVASSOPAKPTVTVLKSKNKEYGLR-LASHIFVKEISQSLSLAARDG | 209 | | | |
| Qy | 297 SRQE--LVERTH-SILREAVAAERQRYWEKKQTLLQFQKSMACQLYREKVNALQAQ | 353 | | | |
| | : - - : : : | : | : : | : : : | |
| Db | 210 NIQEGDVVLKGVINGVTNNMSTLD-----AKTLIERSKGKLMMVORDE----RAT | 255 | | | |
| Qy | 354 VCELOKERDQAYSARDSAQREIS--QSLVEKDSILRRVOVFELTDQVCBLRTLRLQALAEPP | 411 | | | |

```
Db 256 LNVNPDLSSTHSANASRDIDSEIQSLASDHSGRS-----HDP 296
Qy 412 GVLKQEAETPCPCPKQKRLVRMHAIACPRDDSCSLVSTESQL-----LSDL 459
Db 297 -----RRSRSPDQSEPSDHSRHS--PQPSNGSLRDERISKPGAVTPVKHADH 350
Qy 460 SATSREL-VDSFRSSPAPPSQQSLYKRVABDFGEFPWFSFSCLEIPEGDPGALPGAKA 518
Db 351 TPKTVEEVVERNEKQTSLEPKPKVYAAQV-----GNQMWIVLSV----- 390
Qy 519 GDPHLDYELLDATDLPQLLESSLPVSPGRLOVSESGVLMRRPARRILISQVMTLAFQDA 578
Db 391 ---HLMSYL-----IQLMKMGFL---RPSMKLVK-----FRKGD 420
Qy 579 LLEQLSIVGGNLTGIFIHRTVPGSAADQWALRPGQIIVWVDYEAEPFLFKAVLEDTTL-- 636
Db 421 V--GLRLAGNDVGLFVAGVLEDSFAAKEGLEGGQILRVN---NVDFTNIIEAVLFL 475
Qy 637 -----EBAVGLRRVDFGCCLSVKVNTDGYKRLQLDLAKVATSGDSFYIRVNLAMEGR 690
Db 476 LDLPKGEEVTILAOK-----KKDVYRRIVE-----SDVGDGDFYIRTHPEYEKE 518
Qy 691 AKGELQVHCNEVLHVTDTMFQGC--CCWHAHRVNSTYMTKDTAAHGTIPNYSRAQQQLIALI 749
Db 519 SPYGLSFNKGVEFRAVDLTLYNGKLSWLAIRIGKNHKE--VERGIPIPNKRAEQ--LASV 574
Qy 750 ODMTOQCVTTRKPSGGQPKLVIRVSMKAKASPLR--LSFDRGQLDPSRMGSSSTCFWAE 808
Db 575 Q-----YTLPKTAG-----DRADFWRFGLRSSKRNLRKREDLS-----AQ 612
Qy 809 SOLTLVP--YTLWPHRPARPRVLLPVRAVGKILSEKLCLLQGFKKCLAEYLSQEE---Y 864
Db 613 PVQTRFPAVERVLEAGFLRPVTF--GFADVAREK-----LAREEPDIY 657
Qy 865 E-AWSQRGDIQEGEVSGRCWTHAVESLMKETHALLOVLDVSVCTLHRMDIFPIVI 923
Db 658 QIAKSEPRDAGTDQRSSG--YIRLHTTKQIIDQDKHALLDVTPNVADELNYAQRPIVW 714
Qy 924 HYSVNEKM-----AKKIKKGLQRLGTSEEQLEEAARQEEGDLDRAPCLY 967
Db 715 FLNPSKQGVKTMRLCPESRKSARKLYERSHKLAKNNHLLFTT-----IN 762
Qy 968 SSLAPDGNSSDLGLLSCVRQAIADQKKVWVTEQ 1001
Db 763 LNSMNDGW-----YGALKEAVOOQONLWVYSE 790

RESULT 4
A42184
nuclear mitotic apparatus protein NuMA - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 17-Mar-2000
C:Accession: A42184; S33376; S55331; S23647; S24554
R:Compton, D.A.; Szilak, I.; Cleveland, D.W.
J. Cell Biol. 116, 1395-1408, 1992
A:Title: Primary structure of NuMA, an intranuclear protein that defines a novel pathway
A:Reference number: A42184; MUID:92176238
A:Accession: A42184
A:Molecule type: mRNA
A:Residues: 1-2101 <COM>
A:Cross-references: EMBL:Z11584; NID:G35120; PIDN:CAA77670.1; PID:G35121
A:Note: sequence extracted from NCBI backbone (NCBIN:85755, NCBI:85760)
R:Tang, T.K.; Tang, C.C.; Chen, Y.L.; Wu, C.W.
J. Cell Sci. 104, 249-260, 1993
A:Title: Nuclear proteins of the bovine esophageal epithelium. II. The NuMA gene gives
A:Reference number: S33376; MUID:93280231
A:Accession: S33376
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1705-2101 <TAN>
A:Cross-references: EMBL:Z14229; NID:G296118
A:Note: this translation is not annotated in GenBank entry HSNUMAT3G, release 113.0
```

```
R:Harborth, J.; Weber, K.; Osborn, M.
EMBO J. 14, 2447-2460, 1995
A:Title: Epitope mapping and direct visualization of the parallel, in-register array
A:Reference number: S55331; MUID:95300777
A:Accession: S55331
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 247-279 <HAR>
R:Yang, C.H.; Lambie, E.J.; Snyder, M.
J. Cell Biol. 116, 1303-1317, 1992
A:Title: NuMA: an unusually long coiled-coil related protein in the mammalian nucleus
A:Reference number: S23647; MUID:92176231
A:Accession: S23647
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-771 'Q', 773-814, 'ER', 817-872, 'E', 874-1267, 'RLRLQAETASNSARAERSSALREEVQ
A:Cross-references: EMBL:Z11583; NID:G351118; PIDN:CAA77669.1; PID:G35119
A:Note: the authors translated the codon GAG for residue 781 as Gly, TTC for residue
C:Genetics:
A:Gene: GDB:NUMA1; NUMA
A:Cross-references: GDB:137229; OMIM:164009
A:Map position: 11q13-11q13
C:Keywords: mitosis; nucleus

Query Match 5.1%; Score 262; DB 2; Length 2101;
Best Local Similarity 19.9%; Pred. No. 8.7e-06;
Matches 208; Conservative 163; Mismatches 376; Indels 296; Gaps 39;

Qy 4 LCRRSALTALDEETLWEMMESHRRIVRCICPSRITPYLROAKVLCQDDEEVLHSPRL 63
Db 263 LLNKRQAASPLPEKLELRDN-----ESLTMRLHETLKQCQDLKTEKSQMDRK 312
Qy 64 TNSMRAGHLDDLKTRGKNGAIF--LESIKFHPNPDVYTLVTCLOPDVDFSNFSGLMETS 122
Db 313 INQLSE-----ENGDLSPKLRREFASHLQQLQDALNELTEEHKATQEWLEKQA 360
Qy 123 KITECLAGAGISLQELNQEKGKEVLLRRCCQLQHLGLAETRAEGLHGLADHSMRMR 182
Db 361 QLEKELSA--LODKKCLEE--KNEILQKLSQLEHLS-----QLQDNPPQEGK 406
Qy 183 EVSAHFHEVLRLKDEMLSLS-----LHYSNALQEKELAAASRCRSQELLYL-- 228
Db 407 EYLGVDVLQLETLKQEAATLAANNLTQARVEMLETERGQOEAKLLAERGHFEKQOOLS 466
Qy 229 -----LKQELQANVSSCELELQEQSLRT-----ASDQESGDEELNRLK 268
Db 467 LITDLOSSISNLSQAKELEQASQAHGARLTQAQVASTSELTLTNATIQQQ--DOELAGLK 525
Qy 269 EE-----NEKRLSLTFSLAEKDILEQSLDEARGSRQELVERIH 306
Db 526 QOAKEKQAQLAQTLLQQEQASQGLRHQVQLSSSLKQK--EQQLKEV--AEKQEAATRDQH 581
Qy 307 SLREAVAAEROREQYWEKEQTLQFQSKMACQLYRKVNALQAVCELQKEROQAYS 366
Db 582 A--QQLATAAE--ERASLRERDAALKOLE-----ALEKEKAALKEI-----LQOQLQVANE 629
Qy 367 ARDSAQRETSQSILVEKDSLRROVFEL-----TDQVCLRTQLRQLQAEPPG 412
Db 630 ARDSAQTSVTQOREKAELSRVVEELQACVETARQEQHAQVAEELQLRSEQ----- 684
Qy 413 VLKQARTREPCPREKQRLVRMHAIACPRDDSDCSLVSTESQLLSDLSATSSRELVDSPR 472
Db 685 ---QKATEKERYAQEKDQL-----QEQQLAKLSKLVTKSGLEEKR 723
Qy 473 SSSPAPPSQOQSLYKRVADDFGEEPSFSCLEIPEGDPGALPGAKAGDHPHLYELDTAD 532
Db 724 RAADALEEQOQCISLKAEE-----TRSLVEQHKRERKELEERAGRAGLEARLLQGE 776
Qy 533 LPQLESSLPVSPGRLDVSESGVLMRRRPARILISQVMTLAFQGDALLLEQISVIGNLIG 592
Db 777 AHOAETE-----VLRRELAEMAAOHT-----AESECEQL----- 806
```

```

QY 593 IFHRTVPGSAAQOMALRPGQTQIMVDYEAEBPLFKAVLEDTTLEBAVGLLRVDFGFCCL 652
Db 807 -----VKVAAWRDGYE-----DSQOEAEQYAMFOEQ-----LM 836

QY 653 SVKVTNDGKRLQLDLKAVATSGDSFYIRVNLAMEGRAGKELQVHCNEVLHVTDTMFQ 712
Db 837 TLKEBEKARQELQAKKVA--GISHSELQISRONKLAEL--HAN-----880

QY 713 CGCWHARRVNSYTKMTAAHGTIPNYSRAQQOOLIALIQDQCTVTRKPSGGPKLV- 771
Db 881 ----LARAQQOQVEKEVRAQKLADDLSTLQEKMAATSKVARLETLVK--AGEQOETAS 934

QY 772 RIYSMDKAKASPLRLSDFRGOLDPSRME---GSTCEWABSCSLTVLPYTLVWPHRPARPR 828
Db 935 RELVKEPARA-----GDRQPEWLEEQGQFC-----STQA 965

QY 829 PVLVPRVAGKILSEKILCLQGFKKCLAEVLSQEEVEANSQSGDIIIOGEVSGRCWVTR 888
Db 966 ALQAMEREAFQMGNE-----LERURAALESQGOQOBERGQOEREVARLTQERG-----R 1015

QY 889 HAVESLMKNTHALLDVQDLSVCTLHRMDIFPI--VHVSVMNEKMAK-----KLK-----936
Db 1016 AQADLALAKAARAELEMRNLQNALNEQREVFATLQELALAHALTEKEGQOELAKLRGLEAA 1075

QY 937 --KGLQRLGTSSEQLLEAARQEE 957
Db 1076 QIKLEBELRQTVKQLKEOLAKKE 1098

RESULT 5
G02520
plectin - human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 20-Aug-1999
C:Accession: G02520
R:McLean, W.H.I.; Smith, F.J.D.
submitted to the EMBL Data Library, March 1996
A:Reference number: H01385
A:Accession: G02520
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4574 <MCL>
A:Cross-references: EMBL:U53204; NID:g1477645; PIDN:AAB05427.1; PID:g1477646
C:Genetics:
C:Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein S
F:68-283/Domain: alpha-actinin actin-binding domain homology <ACT>

Query Match 4.8%; Score 249; DB 2; Length 4574;
Best Local Similarity 22.4%; Pred. No. 0.00012;
Matches 187; Conservative 112; Mismatches 319; Indels 218; Gaps 34;

QY 135 LQELNQEKQKGVLLRRCCQQLQEHGLGTAETRAEGLHOLEADHSRMKREVSAMFHEVLRL 194
Db 2293 LQEEAEKMKQVAEEAARLSVAQAQAARLQAEEDLAQORALAEKMKLKEKMQAVQAEATRL 2352

QY 195 KDEMLSLSHYSNALQEKELAAASRCRSLQELLYLLKQEL--QRANMVSSCELELQEQ--- 249
Db 2353 KAEALLO-----QKELAQOQARRLQEDKEQMAQQLAEETQGFQRTLEAERQOLEM 2405

QY 250 -----SLRTAS---DQSGDEELNRLKEE---NEKURSLTFSLAEKDILFQSLDEAR 295
Db 2406 SAEARLKLVRVAEMSRQAARAEADAQFRKQAEIEGKLRHTLQELTQEKVTLVQTLQIQR 2465

QY 296 GSQELVELTHSRVAVAAERQYWEKEQTLQFQKSNAC---OLYREKYNALQ 351
Db 2466 QQSDHDAER---LREAIAELEREKEKL--QQEAKLLQLKSEEMQTVOQEQLLQETALQ 2520

QY 352 AQVCE-----LQKER-----DOAYSARDSAQREISQSLVEKDSL-----385
Db 2521 SFLSEKDSLQREFRFTEQEKAKLEQLFQDEVAKAQLRBEQOQOQOQOQOQOQOQOQOQO 2580

```

```

QY 386 --RRQVFELTDQVCELRTOQLQOAEPPGVLKQEAARTREPCPREKQRLVRMHACIPRDDS 443
Db 2581 EARRRQHEAEQVRRKQEEQLQLE-----QORRQOELLAEQNRQLQLEEQH 2632

QY 444 DCSLVSS---TESQLSLSLSSRELVDSSFRSSPAPPQQOQLYKRVAEDFGEEPWSFS 500
Db 2633 RAALAHSEVTSASQVAAATKPLNGRDALDGPAAAEPEHFSFDGLRRKVSAAORLOEAGILS 2692

QY 501 SCLEIPEGDPGALPGAKAGPHLDYELLDTADLPQLESSLQVPSPGRLDVSSEGVLMRRR 560
Db 2693 A-----EELQRLAQGHHTTYDE---LARRE 2713

QY 561 PARRILSQVMTLAFQGDALLEQISVIGNLTGTIHRVTPGSAADQMALRPGTOIVMVDY 620
Db 2714 DVRYHL-----QGR-----SSIALLLKATNEKLSVYAALQROLLSPGTALILEA 2759

QY 621 EAS-----EPLFKAVLEDTTLEAV--GLLRVVDGFCCLSVKVTNDGYK-----662
Db 2760 QAAAGFLLDPVRNRL---TVNEAVKEGVVGPPELHKLKLSAERAVTGYKDPYTCQOISLF 2816

QY 663 RLQD-----LEAKVATSG-----DSFVIRVNLAMEGRAGKELQVHCNEVLHVT 706
Db 2817 QAMQKGLIVREHGIRLEAQIATGGVIDPVHSHRVPVDVAYR---RGYFDEENRVLADP 2873

QY 707 DTMFQCGCGWHARRVNSYTKMTAAHGTIPNYSRAQQOOLIALIQD-MTQOC---TVTRKPS 763
Db 2874 SDDTKGFFDPNTHENTYL-----QLLERCVEDPETGLCLLLPLDKAA 2916

QY 764 SGGPKLIVRIVSMDKAKASPLRLSDFRGOLDP--SRMEGSTCFWAECSCLTVPYTLVWP 821
Db 2917 EGG-----ELVYTD-----SEARDVFEKATVSAFPGKQKTVTIV--EINSEYFT----2961

QY 822 HRPARPRVLLVPRVAGKILSEKILCLQGFKKCLAEVLSQEEVEANSQSGDIIQEG 877
Db 2962 ---AEQRDLRQFRGRTIVEXII-----KIIIVVEQOE-----QKGRLCFEG 3003

RESULT 6
A59404
plectin [imported] - human
C:Species: Homo sapien (man)
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C:Accession: C59404; A59404
R:Liu, C.G.; Maercker, C.; Castanon, M.J.; Hauptmann, R.; Wiche, G.
Proc. Natl. Acad. Sci. U.S.A. 93, 4278-4283, 1996
A:Title: Human plectin: organization of the gene, sequence analysis, and chromosome 1
A:Reference number: C59404; MUID:96210632
A:Accession: C59404
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4684 <STO>
A:Cross-references: GB:CAA91196; NID:g1296662; PIDN:CAA91196.1
C:Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protei

Query Match 4.8%; Score 249; DB 2; Length 4684;
Best Local Similarity 22.4%; Pred. No. 0.00012;
Matches 187; Conservative 112; Mismatches 319; Indels 218; Gaps 34;

QY 135 LQELNQEKQKGVLLRRCCQQLQEHGLGTAETRAEGLHOLEADHSRMKREVSAMFHEVLRL 194
Db 2403 LQEEAEKMKQVAEEAARLSVAQAQAARLQAEEDLAQORALAEKMKLKEKMQAVQAEATRL 2462

QY 195 KDEMLSLSHYSNALQEKELAAASRCRSLQELLYLLKQEL--QRANMVSSCELELQEQ--- 249
Db 2463 KAEALLO-----QKELAQOQARRLQEDKEQMAQQLAEETQGFQRTLEAERQOLEM 2515

QY 250 -----SLRTAS---DQSGDEELNRLKEE---NEKURSLTFSLAEKDILFQSLDEAR 295
Db 2516 SAEARLKLVRVAEMSRQAARAEADAQFRKQAEIEGKLRHTLQELTQEKVTLVQTLQIQR 2575

QY 296 GSQELVELTHSRVAVAAERQYWEKEQTLQFQKSNAC---OLYREKYNALQ 351

```


A: Molecule type: mRNA

A: Residues: 1-1116 <RES>

A: Cross-references: GB:L27476; NID:g498012; PIDN:AAA61300.1; PID:g498013

C: Genetics:

A: Gene: X104

C: Superfamily: guanylate kinase homology; GLGF domain homology

F: 37-116/Domain: GLGF domain homology <GLG1>

F: 731-879/Domain: guanylate kinase homology <GKI>

Query Match 4.8%; Score 246.5; DB 2; Length 1116;
Best Local Similarity 21.2%; Pred. No. 2.5e-05;
Matches 212; Conservative 137; Mismatches 393; Indels 257; Gaps 47;

QY 55 EVVLSPLTNSAMRAGHLLDLKTRGNGAIAFLESLEKHNPDVYTLVTGLQPDVDF 112
DB 94 EDVLHS-FAVQOLKSGKVAIVVPRKQVAAALQA-----SPLDODDRAFEVWDEFG 148
QY 113 SNF-SGLMETSLEKLAGAIGSLOEELNQEKGKEVLLRRCQQLQEHGLAETRAEGLH 171
DB 149 RSFRSGYSERSRLNS-HGGRSRSWED--SPERGRPH---ERARSERDLRDRSRGRSLE 202
QY 172 Q-LEADSRMKREYSAHFHEVRLKDEMLSL-HYSNALQEKELAAASRCRSLEQEEYLL 229
DB 203 RGLDQDHARTD-----RSRGRSLERGLDHDGSPSRDRDRDRSRSID----- 246
QY 230 KQELORANWSSCELELOEQLRTASQESGDEELNKEENKELRLTSLAEKDILEQ 289
DB 247 -QDYERAY-----HRAYPDYER-----AY 265
QY 290 SLDEARGSRQELVERIHSRLERAVAAERQEQWEEKQTLLOFQSKMACQLYREKVN 349
DB 266 SPEYRRCAR-----HDASRG-PRSRREHPHSRSPSPPEGRPGPIGVLLMKSRAN- 316
QY 350 LQAOVCELOERQOAYSARSAQ-----REISQS-LVEKDSLRQ--VFELTDQVCE--LR 400
DB 317 -----BEYGLRGSQIFVKEMTRTGLATKDGNLHEGDIILKINGVTVMNSL 363
QY 401 TOLRQLOAEPGVLQOEARTEPCPREKORLVRMAHATCPDD--SDCSLYSSTES----- 453
DB 364 TDARKLLEKSRGL-QLVVLRD-----SQOTLNIPLNSDSEIEDISEITSEITRSFSPEE 418
QY 454 -QLLSDLSATSRRELVDSPRSSPAPPSSQOOLYKRVAEDFGEPEPWSFSSCLEIPEGDGP 511
DB 419 RRHQYSDYDHSSSEKLEKRPSSREDTPSRLS-----RMGATPTPFKS-----TG 463
QY 512 ALPGAKAGDPHLYELDTADLPQLESSLPVSPGLDVSESVLMRRRPARRIL--SQV 569
DB 464 DIAGTVVPETNKPEYQOEPAP-----QPKAAPR-----TFLRSPDEAIYCPNT 510
QY 570 TMLAF-QGDALLEQISVIGNLGTGIFHRTVTPGSAADQMALRPGTQIVMVDYEAASEPLFK 628
DB 511 KMYRFKKGDSV--GLRAGNDVGVIFVAGIOEGTSAEQEGLQEGDQILKVNQDFRGL-- 566
QY 629 AVLEDITLLEAVGLLRVVG-FCLSVKVNVDGYKRLQDLQAKVATSGDSFYIRVNLAM 687
DB 567 -VREDAVLY-----LLEIPKGMVYTLIAQSRADYRDIL-----ACGRGDSFFIRSHFEC 615
QY 688 EGRAKELQVHCNEVLHVTDTMFOG-CGCWHAHRVNSYTMKDTAAHGTIPNYSRAOQOLI 746
DB 616 EKETPOSIAFTGCEVRVVDVTLTDGKLGWLAVRIGNELEK-----GLIPNKSRAEQ--M 668
QY 747 ALTQDMTOQCTVTRKPSGGPKOKLRIVGMDKAKASPLRLSLFSDRGOLDPSRMEGSCTFW 806
DB 669 ASYQN-----AQRNAGDADRDFWRM-----RGQRSGVKKNLRKSRDLTAVVSVSTKFP 717
QY 807 AESCLTLVPYTLVWPHRPARPRVLLPRAVAGKILSEKLCILQGGFKCLAEYLSQEEYEA 866
DB 718 A-----YERVLLREAGFKRPVILF-GPIADIAMEKLA-----NELPD 753
QY 867 WSQRGDIQEGEYSGRCWTR-HAVESLMENKNTIALDLVDQDSVCTLHRMDFIPTVIHV 925
DB 754 WFOTAKTEPKDAGSEKSTGVVRLNTVQVIEODKHALLDVTPKAVDLLNTQWFSVIVSF 813

QY 926 SVNEKM-----AKKLKKGRLQRLGTSEEQLLEAARQEGDLDR 962
DB 814 TPDSRGQVNTMRQRLDPTSNSSRKLFKDHANKLKCAHLFTATINL----- 860
QY 963 APCLYSSLPADGWSDDLGLLLSCVROAIADBOKKVWVTEQ 1001
DB 861 -----NSANDSW-----FGSLKDTIQHOQGEAVVWSE 887

RESULT 9

T47237

myosin II heavy chain [imported] - Naegleria fowleri (fragment)

C: Species: Naegleria fowleri

C: Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C: Accession: T47237

R: Shaw, D.R.; Sullivan, P.K.; Marciano-Cabral, F.; Ennis, H.L.

submitted to the EMBL Data Library, December 1995

A: Description: Codon usage in Naegleria fowleri.

A: Reference number: Z24413

A: Accession: T47237

A: Status: preliminary; translated from GB/EMBL/DBDJ

A: Molecule type: mRNA

A: Residues: 1-746 <SHA>

A: Cross-references: EMBL:U43192; PIDN:AA01786.1

A: Experimental source: strain LEE mp; cell type amoeba

Query Match

Best Local Similarity 4.8%; Score 245; DB 2; Length 746;

Matches 160; Conservative 151; Mismatches 293; Indels 228; Gaps 31;

QY 47 KVLQCDDEEVLSPLTNSAMRAGHLLDLKTRGNGAIAFLESLEKHNPDVYTLVTGL 106
DB 19 KKLSESSKDELTQLNKTND--EKNELVNKLKKAED-----LKNLKKSKDD-----L 64
QY 107 QPVDVPSN-----FSGLMETSLEKLAGAIGSLOEELNQEKGK--EVLV----- 150
DB 65 QAEKDDSDNRIRKLEODLREKEQLSENLAIRIADLENEARTKEAOKKSTEMELSSVKDDL 124
QY 151 ----RRCQOLOEHLGLAETRAEGLHOLEADHSRMRKREVSAPHFHEVRLKDEMLSLSHVS 206
DB 125 NRTKQRAEQLOSDLEAQRERANELENLLSDTEGGKNQLDSQFKQL-----Q 170
QY 207 NALQEKELAAASRCRSLOEELYLLKLOELRANWSSCELELOEQLRTASQESGDEL-N 265
DB 171 NELQNETNLQMKSENERL-----ORELEEMK-RSLSDKQNESTSLDS 213
QY 266 RLKEENEKLSLTFSLAEKDILEQSLDEARGSRQELVERIHSL---RERAVAAERQEQY 322
DB 214 KVKSLKDKIRELTALLETSSRTDLKKRKMDEKVRKLAQLOEQALQGETQKK-- 271
QY 323 WEEKEQTLQFOKSKMACQLYREKVN-----ALQAOVCBL 357
DB 272 -NDADNRVKOLESELQGVKSRDLNKLNTSGDMNGLKRLQDSENNLVAKLAEIQKL 330
QY 358 QKEROQAYSARSAQBEISQSLVEKDSLRQVFTLDQVCELRQLRQ-----LQAP 410
DB 331 QKDLSDHGDRETEBEQL-----DALRKLQELTSLSDANQKTQOEASRQNLSEN 383
QY 411 PGVLQOEARTPCPREKORLVR-MHAICPRDDSDCSLSVSTESOLLSDLSATSSRELVD 469
DB 384 NRLKSEVSLREDLQNERPRKQEMERVQSENE-----KSELLTQLQ----- 427
QY 470 SFRSSPAPPSSQOOLYKRVAEDFGEPEPWSFSSCLEIPEGDPALPGAKAGDPHLYELID 529
DB 428 -----KLQEAYSEVKDELKD-----LSKNASRGSGVGVGSDSAEVEKLRREYEM-- 471
QY 530 TADLPQLESSLPVSPGLDVY-----ESGVLMMRRRRPARRILSOVTHMLAFQ 576
DB 472 --QLAQLKARVEEVTPQQRVDVENKRSRVEDMTKMTLQTEERLKRKVEQOKKSVEMEC 529
QY 577 DALLEQISVIGNLGTGIFHRTVTPGSAADQMALRPGTQIVMVDYEA-SEPLFKAVLED-- 633

Db 530 DELRE-----LAEEADLRDELNRTKLEHOALIQQLRDLQERH 569
QY 634 --TTLEEAVGLLRVDGFCCLSVKNTDGYKRLLODLEAKVATSGDSFVIRYNLAMEGRA 691
Db 570 SPASAEENATQKR-----EIEEQDLEQERA-----KLDEAA 603
QY 692 KGELOVHCNEVLHVDTTVPQGGCWHHRVNSYTKMD-TAAHGTTIPNYSRAQQOIALIQ 750
Db 604 RRLKQOYENEILDNLNQAQ-----AKERSAASRDMKKADRLREYQRRFEQEARAQ 657
QY 751 DMTQCTVTRKSSGGPKLVIRVSMRKAASPLRLSFDGQDPSRMGSS 802
Db 658 DLEQRUTKVEREN-----KLQ--SOSOSDAS-----KYQAEQEKORLEAEN 698

RESULT 10
T30171
ninein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30171
R:Bouckson-Castaign, V.; Moudjou, M.; Ferguson, D.J.P.; Mucklow, M.; Belkaid, Y.; Milon,
J. Cell Sci. 109, 179-190, 1996
A:Title: Molecular characterisation of ninein, a new coiled-coil protein of the centrosome
A:Reference number: 220751; MUID:96431720
A:Accession: T30171
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-2168 <BOU>
A:Cross-references: EMBL:U040342; NID:g1113864; PID:g1113865; PIDN:AAA83234.1
A:Experimental source: strain C57Bl/6
A>Note: localised specifically in the pericentriolar matrix of the centrosome

Query Match 4.7%; Score 243.5; DB 2; Length 2168;
Best Local Similarity 19.3%; Pred. No. 8.6e-05;
Matches 246; Conservative 183; Mismatches 380; Indels 443; Gaps 61;

QY 13 ALDEETLWEMSHRHVRICPCSLRTPYLRQAKVLCODEEVLHSPRLNSAMRACH 72
Db 639 STEALVTEQMEQHH--RDLCRLRL-----ELEKVRHYEQLDQDTRVASQ 584
QY 73 LIDLKTRKNGAIAFLESKFHNPDVYTL---VTGLQPDV-DFSNFSGLM-----119
Db 685 EQAAMKQKVEQG-----VHTLEKRVSELSEIADLEQAAVLEAHKASC 730
QY 120 ----ETSKLTELAGAIGSLQELNQEKGQKVEILLRCCQL-----QEHGLA-----ET 165
Db 731 RHEEKRLQMAFDEKAQQLBELRQE--HEREQLARQQAATFRQEREGLAQAAWTEE 788
QY 166 RAEGHLQ-----LEADHSRMKREVSARHFVRLKDEMLSLHYSNALQE-KELAA 216
Db 789 KYRGLEQSQVQEQQLSLSEKHALEKEE-----LREE---LSEHRRLEQEGREME 835
QY 217 SCRSLQEBELYLLKQELORANMVSCE--LELQEOSLRTASDOESGDEELNRLKEENKL 274
Db 836 TECNRRVSQI-----EAQCOADCEKVTEHCQTLQSLVRHR--QELRLDLDQHL 885
QY 275 RS-LTFSLAEKDILQSLDEARGSRQELVERTHSLRERAVAAERQEQ-----TWE---324
Db 886 RSQWEFF--EKDELAQECTDAQEQLEKALQ-----RERATAAMKQEQEILERTYKDLRN 937
QY 325 ----EKEOTLLQFQSKMACQLYREKVNALQAOVCELOKEROQAYSARSAQ-----372
Db 938 ILUSTEREQLLDLKLQNASQ---SOHGLLSQIILEL--KRSQERELRQGGALCQTGVS 992
QY 373 -----REISQSLVEKDSLRRQVFELTDQ-----VCLELRTQLRQ 405
Db 993 EQLASPELERLRVEHEQEREMTGKLAALSAHRAHSLERADQEKAEEMSTEICRLQNTQV 1052
QY 406 LQ-----AEPGVL-----KQEA 418
:

Db 1053 MQAAALLMLOGGCOQATAGEAEAGDGAMSLLOQGEQLLEENGDLVLSLQRAHEHAVKENA 1112
QY 419 RTREPCPREKQRLVRMHAICPRDDSCSLVSVSTESOLLSDLSATSSRELVDVSRSSSPAP 478
Db 1113 KWATEIYQOQRLLKLI-----EPGSVISSCLEGTSEISG--SSREQVEPTMKOGPAT 1163
QY 479 PSQOSLYKRVAEDEGPEPWSFSCLEIPEIGDPGALPG-----AKAGDPHLD-YEL 527
Db 1164 -----KHFLSDLGDH-----EARDLASTGSTSVQKCKTEASEASLDCFSE 1205
QY 528 LOTADLPOLLESSLOPVSPGRGLDVSEGLVMRRRPARILISQVTLMAFOGDA-----LL 580
Db 1206 LENSEDTRTESDL-----KSIQSOLREQLTVLRADCCDRASERKODLL 1247
QY 581 EQISVIGGNLTGIFIRHVTGSG-----AADQMALRPGTQIYVMVDYEASEPLFKAV 630
Db 1248 FDISVLKKKLK--MLERLPEASSRYKVLVEDAARENSCLOEELRLVETREESLDSNKL 1305
QY 631 -----LED-----TTLEAVGLLRVDGFCCLSVKVNFTDGYKRLLODLEAKV-----672
Db 1306 TAEVYRLQDEMKKMEVEMETFLSLEKSYD-----EVKVENEELRALVLRLOGKMEKVLGR 1360
QY 673 -ATSGDSFVIRYNLAMEGRAKGE-LOVHCNEVL-----HVTDTMFQ 712
Db 1361 AALQGDSY-----ASGKPPPSSENLEVASDEKMLELRQTPKECTPKVYSMMHHIEBCTOE 1413
QY 713 CGC-----WHAHRVNSYTKMDTAAHGTTIPNYSRAQQOIALIOD--M 752
Db 1414 TQCEGSGTKLLARIKAHEIAFWHRAIKTHPEPSQNRVPEGSAA---LLGL-QDKHL 1469
QY 753 TQCTVTRKSSGGPKL-----VRIVSMDKAKASPLRLSFDGQD---PSRMEGS 801
Db 1470 QEATIAELELE--KQKLQELTRNLREVRTALVRQXDAP-----SQQKEEELKAMMQPA 1522
QY 802 STCFWAESCLTLPYTLVMPHPPARPVLLVPRAVKILSEKLLCLO-----GF 851
Db 1523 VTCGEMQRKVELLYE-----SEKLEENSILRNEITTLNEEDSI 1562
QY 852 KKCLAEYLSQEEYEAWSQRGDIQEGEVSGGRCWVTRHAVESLMKENTHALLDQDLSVC 911
Db 1563 SNLKLBEELNGSQEELWQKTIETIQE-----KASIQTWVKLKKQVSDLKIKN--1609
QY 912 TLHRMDIFPIVHVSNYKMAKKKGLQRLGTSEBOLLE--AARQEEG-----958
Db 1610 --QQLD-----SENIELSQNQNKEKLTNQLAEMLCQREBPACTSEKWEQEN 1659
QY 959 ----DLDRAPCLYSSILAPDQWSDLDGLLSCVR-QAIADEQKKVWVTEQSPR 1004
Db 1660 ASLKEELDHVKYQVOTSTLV-----SSLEAELSRIKLQTHVMEQENLLKDELER 1707

RESULT 11
152300
giantin - human
N:Alternate names: gcp372
C:Species: Homo sapiens (man)
C:Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 20-Aug-1999
C:Accession: I52300
R:Sohda, M.; Misumi, Y.; Fujiwara, T.; Nishiohka, M.; Ikehara, Y.
Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994
A:Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized
A:Reference number: I52300; MUID:95100974
A:Accession: I52300
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-3225 <RES>
A:Cross-references: GB:D25542; NID:g662389; PIDN:BAA05025.1; PID:g808869
C:Superfamily: giantin

Query Match 4.7%; Score 242.5; DB 2; Length 3225;
Best Local Similarity 19.3%; Pred. No. 0.00017;
Matches 220; Conservative 194; Mismatches 396; Indels 331; Gaps 49;

| | | | |
|----|------|--|------|
| Qy | 484 | LYKRVADGEPEPWSFSSC-----LEIPGDPGALPGAK-----AGOPHLDYELLDT | 530 |
| Db | 1211 | DYNRLQFQDEQKENENIGDQLRLQIQVRESIDGKLPSTQDQESCSSTPGLEEPLFKA | 1270 |
| Qy | 531 | AD-----LPOLLESSLQVPSP-----GRLDVSESGLVMRRRPARRI-----LS | 567 |
| Db | 1271 | TEOHHTQPVLESNLCPDWFSHSEDASALOGGTSVAQIKAKLKEATEAKVELELKVSTTS | 1330 |
| Qy | 568 | QVTMLAFQGDALLEQISVIGGNLTGIFTHRVTPGSAADOMALRGQTQIVMVDYEAESPFL | 627 |
| Db | 1331 | ELTKKSEEVFLQEQI-----NKQGLETESL-----KTVSHEAEVHAESL | 1370 |
| Qy | 628 | KAVLEDTTLEEAVGL--LRRVDGFCCLSVKYNTDGYKLLQDLQLEAKVATSGDSFYRYNL | 685 |
| Db | 1371 | QOKLESQLOIA-GLEHLR-----ELQPKLDELQKLTSKKEEDVS-----YLSGQL | 1415 |
| Qy | 686 | AMEGRAKGELQVHCNVLHVTDTPMGQGCWHRAVRNSYTMKDTAAHGTTIPNYSRAOQOL | 745 |
| Db | 1416 | SEREAAALTKIQ---TEIIQEGLI-----KALHTQLEMQAQAKEDER | 1453 |
| Qy | 746 | IALIOWTQOCTVTRPSSGGPKLVRIVSMDKAKASPLRSPRGOLDPSRMEGSGSTCF | 805 |
| Db | 1454 | IKOLO--VELCEMKQKPEEIGESRAKQIQIRKLAALI-----SRKEALKE-- | 1498 |
| Qy | 806 | WAESCLTPVPTYLVMPHRPARPRPVLLVPRAVGKLTLSKLCLOGFKKCLAEVLSQEEYE | 865 |
| Db | 1499 | -----NKSQLELSLNGTTERUTKSIAOVESQ | 1526 |
| Qy | 866 | AWSQRGDIHQEGVSGRCWVTRHAVESLMENKTNHALLDVQ-LDSVCTLHRMDIFPTVTH | 924 |
| Db | 1527 | VSQAQ---KEKDTVGLRLLAQEERDKLITEMDRSLLENQSLSSCSLKLALGL--- | 1579 |
| Qy | 925 | VSVNERMAKMLK-----KGLQR-----LGTSE-----EQLELEAA | 953 |
| Db | 1580 | TEDKEKLVREIESLKSSTAESTEWEQKHKELOKEYEILLQSYENVSNAEARIQHVVYEA | 1639 |
| Qy | 954 | ROEGDLDRAPCLYSSL-----APDGSWLDGLLSCVCRQAIADQEKVYVWTE | 1000 |
| Db | 1640 | RQEKQE-----LYGKLIRSTEANKETEKQJQEAQBEQMEEMKMRKFAKSKQOKILELE | 1693 |
| Qy | 1001 | Q 1001 | |
| Db | 1694 | E 1694 | |

RESULT 13
JE0366
tight junction protein, ZO-2 - chicken
C:Species: Gallus gallus (chicken)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 21-Jul-2000
C:Accession: JE0366
B:Collins, J.R.; Rizzolo, L.J.
Biochem. Biophys. Res. Commun. 252, 617-622, 1998
A:Title: Protein-binding domains of the tight junction protein, ZO-2, are h
A:Reference number: JE0366; MUID:99057550
A:Accession: JE0366
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1163 <COL>
A:Cross-references: GB:AF085184; NID:G3820579; PIDN:AAAC95469.1; PID:G3820580

```

Query Match      4.7%; Score 242; DB 2; Length 1163;
Best Local Similarity 22.2%; Pred. No. 4.5e-05;
Matches 147; Conservative 102; Mismatches 239; Indels 174; Gaps 30;

Qy 376 SOSLVEKDSLRQVFELTDQVCELTQLRQLQAEPPGVVLKQEARTEPCPREKQLVRMH 435
      ::::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 333 ARKUIERSGKLQVLWLRD-----KQFLNIP 360

Qy 436 AICRPDSDCSL--VSTESOLL-----SDLSATSSRELVDTSRSPAPPQQS 483
      ::::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 361 SL----NPSDSMDISETESNRSFSDRLRHSDLHSSNEKLKKTNAADDPNRM 417

```

| | | | |
|---|------|--|-----|
| Qy | 484 | LYKRVADFGEPEPMSFSCSLEIPEGDPGALPGAKAGDPHLDYELLTDADLPQLLESSLPQV | 343 |
| Db | 418 | -----RMGAMPTPFKSTGDI-----ATTAVTVVDANKEPKYQDDVAVPOPKAVTRII | 464 |
| Qy | 544 | -SPGRLOVDSGCVLMRRRPARRILISQVTMLAF--QGDALLEQIISVIGCNLTGFIHRVTG | 601 |
| Db | 465 | LKPSPEDEAIG-----PNTKMYRFKKGDSV--GLRAGLNDGVIPIAGIQBG | 510 |
| Qy | 602 | SAADQMALRPGCTQIVMDVYEAASEPLFKAVLIEDTTLLEAVGLLRV---DGFCCLSVKVMT | 658 |
| Db | 511 | TSADQEGLGQEDQILKVNQD---FRGIYR---EDAVLYLLEIPKGD-----TVTILA | 557 |
| Qy | 659 | DGYKRLLODLEAKVATSGDSFYIRVNLAMGRAGKELQVHCNEVLHVDTMTFQG--CGCWH | 717 |
| Db | 558 | QSKYEYVRDIMA--CGRGDSFFIRSHFPECEKESQSLAFTRGEIFRVVDTLYDGKLGNL | 615 |
| Qy | 718 | AHRVNSVTMTKDTAAHGTIPNVSRQAQOLIALIODMTQCTVTRKPSGGPKOLVRIVSMD | 777 |
| Db | 616 | AVRICNELEK-----GLIPNRSRAEQ--MASVQNAQD-----GPS-----D | 850 |
| Qy | 778 | KA-----KASPLRLSFRDQGLDPSRMEGSSCTCFWAESCLTLVPTLVMPHPRPRPV | 830 |
| Db | 651 | RADFWRTRGQSGVKKLNLRKSREDLTAIVSGTKFPA-----YERVQLREAGFKRPV | 702 |
| Qy | 831 | LLVPRVAGKILSEKLC-----LLOGPKKCLAEYLSQEEYEAMQSQRGDIIOEGEVS | 885 |
| Db | 703 | VIF-GPIADVAMEKLSLTDLPHLVYOTAK-----TEPRDAGSEKATGV----- | 742 |
| Qy | 886 | VTRIAVESLMEKNTHALLDVQLDQSVCTLHRMDFIPPIVHVSVNEKMAKKLKGQLRL | 942 |
| Db | 743 | VRLTVTRQIIEQDKHALLDVTPRAVDLLNNTQWFPIVVFNFNPSKOGVKTMR--QRL | 800 |
| Qy | 943 | -GTSEQLLEAARQEGDLDR--APCLYSSLAIDPGWSDLDGLLSCVQRQIADEBQKKV | 999 |
| Db | 801 | SNKSRKLYEQANKLKKTCSHLETTATINLSANDSW-----YGSCLKDTIQOQGE | 854 |
| Qy | 1000 | EQ 1001 | |
| Db | 855 | SE 856 | |
| RESULT 14 | | | |
| I46236 | | | |
| tight junction protein - dog (fragment) | | | |
| C:Species: Canis lupus familiaris (dog) | | | |
| C:Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 16-Dec-1999 | | | |
| C:Accession: I46236 | | | |
| J. Cell Biol. 124, 949-961, 1994 | | | |
| R:Jesaitis, L.A.; Goodenough, D.A. | | | |
| A:Title: Molecular characterization and tissue distribution of ZO-2, a tight junction protein | | | |
| A:Reference number: A54475; MUID: 94179414 | | | |
| A:Accession: I46236 | | | |
| A>Status: preliminary; translated from GB/EMBL/DBJ | | | |
| A:Molecule type: mRNA | | | |
| A:Residues: 1-775 <JES> | | | |
| A:Cross-references: GB:L27152; NID:g463047; PID:g507892 | | | |
| C:Genetics: | | | |
| A:Gene: ZO-2 | | | |
| C:Superfamily: guanylate kinase homology; GLGF domain homology | | | |
| F:101-171/Domain: GLGF domain homology <GLG3> | | | |
| F:316-464/Domain: guanylate kinase homology <GKI> | | | |

```

Query Match      4.6%  Score 236.5;  DB 2;  Length 775;
Best Local Similarity 23.5%;  Pred. No. 5.1e-05;
Matches 133;  Conservative 86;  Mismatches 231;  Indels 115;  Gaps 26;

Qy 454 QLLSDLSATSSRELVDLFSRSSPAPPQQSLYKVAEDFGEEPWFSSCSLEIPEGDPCAL 513
      ||| ||| :
Db 6 QQYSDYDVHSSNEKL-----KRPNSREDQNRWSR--MGATPTTPFKSM-----GDTASV 53
      :|:
Qy 514 PGAK-AGDPPLHLYELLDTADLPOLLESSLPQVSPGRGLDVSSGVLMMRRRRPARRIL--SOVTV 570

```

Db 54 VGTENKEPRYQEE-----PPAPQKAAPR-----TFLRSPDEDAIYGNTK 96
QY 571 MLAF-QCDALLEQISVIGNLGTIFHRVTPGSAADQMALRPGTQIVMVDYEAASEPLFKA 629
Db 97 MVRFKGDSV--GLRAGGNDVGIFVAGIOEGTSAOEGLOEGDQILKVTQDFRGL--- 151
QY 630 VLEDTTLAEAVGLRRVDG--FCCLSVKVNVDGYKRLQLDLEAKVATGSDSFYIRVNLAME 688
Db 152 VREDAVLY-----LLETPKGMVITLAQSRADVYRDIL-----ACGRGDSFFIRSHPECE 201
QY 689 GRAKGELOVHCNEVLVHTDPMFOG-CCCHAHVRVNSVTMKDTHAAGTIPNYSRAQQOLIA 747
Db 202 KETPQSLAFSRGVEFRVVDLYDCKLGHVLAVRIGNELEK-----GLIPNKSRAEQ--MA 254
QY 748 LIQDMTOOCTVTRKPSGPGQKLVRIVSMDKAKAPLRLSFDRGQLDPPSRMEGSSCFWA 807
Db 255 SVQN-----AQDNAGDRADFWM-----RGQSRGMKNLRSREDLTAASVSVTKFPA 303
QY 808 ESCLTLVPYTLVWHPARPRLVLL-----VPRAVGKILSEKCLCQLQGFKKCLAEYLSQEE 863
Db 304 -----YERVLRLREAGFRKRPVVLFGPIADIALEKLANELPDLFTQAK-----TEP 347
QY 864 YEWSORGDIIOEGEYSGGRCWTRHAVESLMEKNTHALLDVQDLSVCTLHRMDIPRVI 923
Db 348 KDAGSEKSSGV-----VLNVTROIIEODKHALLDVTPKAVDLLNYTQWPPIVI 396
QY 924 HVSVNEKMAKKLKGQLRGLTSEQLLEAARQEBGDLDRAPCLY-----SSLAPDGWS 976
Db 397 FFPDSRQGVKTKR--QRLNPTSNKSRKLYDQANKLKT-CAHLFTATINLSANDSW- 452
QY 977 DLGGLSCVRQAATADEOKKVVWTEQ 1001
Db 453 -----FGSLKDTIQHQOGEAVVWSE 472

RESULT 15
A33977
myosin heavy chain, nonmuscle - chicken
N:Contains: myosin ATPase (EC 3.6.1.32)
C:Species: Gallus gallus (chicken)
C>Date: 31-Dec-1993 #sequence.revision 31-Dec-1993 #text_change 19-Jan-2001
R:Shonet, R.V.; Conti, M.A.; Kawamoto, S.; Preston, Y.A.; Brill, D.A.; Adelstein, R.S.
Proc. Natl. Acad. Sci. U.S.A. 86, 7726-7730, 1989
A:Title: Cloning of the cDNA encoding the myosin heavy chain of a vertebrate cellular my
A:Reference number: A33977; MUID:90046668
A:Accession: A33977
A:Molecule type: mRNA
A:Residues: 1-1959 <SHO>
A:Cross-references: GB:M26510; NID:g212382; PIDN:AAA48974.1; PID:g212383
R:Katsuragawa, Y.; Yanagisawa, M.; Inoue, A.; Masaki, T.
Eur. J. Biochem. 184, 611-616, 1989
A:Title: Two distinct nonmuscle myosin-heavy-chain mRNAs are differentially expressed in
s.
A:Reference number: S06116; MUID:90032648
A:Accession: S06116
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 716-1008 <KAT>
A:Cross-references: GB:X17589
R:Hodge, T.P.; Cross, R.; Kendrick-Jones, J.
J. Cell Biol. 118, 1085-1095, 1992
A:Title: Role of the COOH-terminal nonhelical tailpiece in the assembly of a vertebrate
A:Reference number: A43422; MUID:92381096
A:Accession: A43422
A>Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1900-1959 <HOD>
A:Experimental source: brush border
A:Note: sequence extracted from NCBI backbone (NCBIP:111947)
C:Superfamily: myosin heavy chain; myosin motor domain homology

C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleo
F:84-764/Domain: myosin motor domain homology <MMOT>
F:174-181/Region: nucleotide-binding motif A (P-loop)
F:552-565/Region: actin binding #status predicted
F:626-640/Region: actin binding #status predicted
F:837-1936/Domain: coiled coil #status predicted <COI>
F:837-1277/Region: S2
F:1278-1959/Region: light meromyosin
F:1937-1959/Domain: carboxyl-terminal <CBT>
F:125/Modified site: N6, N6-trimethyllysine (Lys) #status predicted
F:180/Binding site: ATP (Lys) #status predicted
F:694,704/Active site: Cys #status predicted

Query Match 4.6%; Score 235; DB 1; Length 1959;
Best Local Similarity 20.8%; Pred. No. 0.00021;
Matches 244; Conservative 170; Mismatches 385; Indels 376; Gaps 58;

QY 14 LDEETLWEMMESHRH---RIVRCICPS-----RLPYLRQAKVLCOLDDEEVLHSPR-- 62
Db 649 LYKEQLAKLAKMATLNTNPFRVCIIPNHEKKAGKLDPHL-----VLDQLRGCVLEGIRIC 704
QY 63 -----LTNSAMRAGHLDDLKTRGKNGAIAFLSLEFHPDVVTL- 102
Db 705 RQGFNNRVVFEFRORYEILTPNAPKGF-----DGKQACVLMIMAKLEL-DSNLRIG 757
QY 103 -----VTGLQPDVDFSNFSGMLMETSCLTECLAG-----AIGSLQELN 140
Db 758 QSKVFFRAGVLAHLEERDL-----KITDVIIGFOACCGCYLAKFAKARQOOLT 807
QY 141 QEKGQEVLLRRCQOQLEHGLAE-----TRAEGHLEADHSRMRKREVSAMFHEVLR 193
Db 808 AMK-----VLQRNC---AAYLKLNRNWMWRLFTVKVPLLQV-----SROEEEMMAKEELIK 856
QY 194 LKDEMLSLSHYSNALQEKELASRCRSLOEELYLLKQELQANMVSSCELEQEQSLR- 252
Db 857 VKEKQLAA-----ENRLESEM-----TFQQLMAEKMLQOE-QLQAEALCAEAEIRA 904
QY 253 --TASQESGDEEL-----NRLKEENKLSLTFSLAEK-----DILQSLEADARGSR 298
Db 905 RLTAKKQEL--EETCHDLEARVEEERECQHLQ---AEKKKMQONIQELEEQLEEEESAR 959
QY 299 QEL-----VERIHSLRERAVAAERQEQWEEKEOTLLQFKQSKMACOLYREK----- 346
Db 960 QKIQLEKVTTEAKLKLKEEDVIVLEQNLKL--AKEKKLLEDRMSEPTTLNTEEEKSKS 1017
QY 347 -----VNALQAVCELOKQERDQAYSARDSAQR---ETISQSLVEKDS--LRRQVFELTQVC 397
Db 1018 LAKLKNKHEAMITDLEER-----LRREKQROELEKTRKLEGSSDLHDQIAELQAOIA 1072
QY 398 ELRTQRLQQAEPGVLKQEARTEPCPREKQRLVRMHAICPRDDSDCSLVSVSTESQLLS 457
Db 1073 ELKIQLSKKEELQAAAL---ARVEEAAQKNMALKKIREL-----ESQI-- 1113
QY 458 DLSATSSRELVDSSFRSSPPAPQSOOSLYKRVADPGEPPWSFSSCLPEIGDPGALPGAK 517
Db 1114 ----TELQEDLESERASRNKAEKQK-----RDJGELEALKTELEDTTLDSTAAQOELR 1162
QY 518 AGDPH-----LDYELDTADL--PQLESSLQVPSPGRLDVSESGVLMRRRPA----- 562
Db 1163 SKREQEVTVLKKTLDEDEAKTHAQIQEMRKQHSQAIEELAEQLEQTKRVKANLEKAKOAL 1222
QY 563 ---RRILSQVTMLAFQ-----DALLQOISVIGNLGTIFHRVTPGSAADQMA 608
Db 1223 ESERAEIISNEKVLLQKGDAEHKRVKVDQLOQLQV-----KFTGE----- 1265
QY 609 LRPTQIVMVDYEAASEPLFKAVLEDTTLEEAVALLRVRVDGFCCLSVKVNVD--GYKRLQ 666
Db 1266 -----RVKTELAERVNKLQVE---LDNVTGLLNOSDS---KSIKLANDFSALQSLO 1311
QY 667 DLEAKV-----ATSGDSFYIRVNLAMEGRAGELQ-----VHCNEVLHVD 707
Db 1312 DTQELLQETRLKLSFSTKLKQTEDEKNALKEQLEEEEAARNLEKQISVLLQQAQVAREK 1371

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2002, 09:19:57 ; Search time 39.24 Seconds
(without alignments)
2841.949 Million cell updates/sec

Title: US-09-767-215-2

Perfect score: 5149

Sequence: 1 MGELCRRDSALTADEETLW.....VROAIADEQKVVWTEQSPR 1004

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|--------------------|
| 1 | 5149 | 100.0 | 1004 | AAE07164 | Human caspase recr |
| 2 | 5034 | 97.8 | 1139 | AAE07165 | Human predicted ca |
| 3 | 1239.5 | 24.1 | 1147 | AAU01207 | Human caspase recr |
| 4 | 887.5 | 17.2 | 1032 | AAU01206 | Human caspase recr |
| 5 | 566.5 | 11.0 | 746 | RAM93822 | Human polypeptide, |
| 6 | 512.5 | 10.0 | 536 | AAU01204 | Rat caspase recrui |
| 7 | 493 | 9.6 | 536 | AAU01205 | Human caspase recr |
| 8 | 396.5 | 7.7 | 366 | AAU01205 | Human protein sequ |
| 9 | 309 | 6.0 | 350 | AAU01205 | Human polypeptide |
| 10 | 282 | 5.5 | 1445 | ABB64619 | Drosophila melanog |
| 11 | 262 | 5.1 | 2101 | AAU01205 | Sequence of the in |

| | | | | | | |
|----|-------|-----|------|----|----------|--------------------|
| 12 | 262 | 5.1 | 2101 | 22 | AAU01205 | Nuclear mitotic ap |
| 13 | 262 | 5.1 | 2207 | 22 | AAU01205 | Novel human secret |
| 14 | 259 | 5.0 | 2101 | 21 | AAU01205 | Human NUMA protein |
| 15 | 258 | 5.0 | 1916 | 22 | ABB62423 | Drosophila melanog |
| 16 | 258 | 5.0 | 1316 | 22 | ABB66089 | Drosophila melanog |
| 17 | 254.5 | 4.9 | 2246 | 22 | ABG05850 | Novel human diago |
| 18 | 252 | 4.9 | 2117 | 22 | AAU01204 | Novel human secret |
| 19 | 252 | 4.9 | 2192 | 18 | AAU01204 | LexA/NUMA fusion p |
| 20 | 252 | 4.9 | 2272 | 18 | AAU01204 | GAL4/HA/NUMA fusio |
| 21 | 248.5 | 4.8 | 2442 | 21 | AAU01204 | Human cytoskeletal |
| 22 | 246.5 | 4.8 | 2115 | 21 | AAU01204 | Human NUMA protein |
| 23 | 240.5 | 4.7 | 1960 | 22 | AAU01204 | Human protein sequ |
| 24 | 238 | 4.6 | 3248 | 17 | AAU01204 | Kinetochore protei |
| 25 | 237 | 4.6 | 2482 | 16 | AAU01204 | Human mitotin. Ho |
| 26 | 237 | 4.6 | 2482 | 19 | AAU01204 | Human mitotin. Ho |
| 27 | 236.5 | 4.6 | 2096 | 21 | AAU01204 | Human ORFX ORF1356 |
| 28 | 236.5 | 4.6 | 2139 | 22 | AAU01204 | PN7771. Homo sapi |
| 29 | 235 | 4.6 | 612 | 22 | AAU01204 | Human protein sequ |
| 30 | 233.5 | 4.5 | 1456 | 22 | AAU01204 | Drosophila melanog |
| 31 | 233 | 4.5 | 1963 | 22 | AAU01204 | Human protein sequ |
| 32 | 232.5 | 4.5 | 1037 | 22 | AAU01204 | Novel human diago |
| 33 | 230 | 4.5 | 931 | 22 | AAU01204 | Novel human diago |
| 34 | 229.5 | 4.5 | 2415 | 22 | AAU01204 | Novel human diago |
| 35 | 228 | 4.4 | 674 | 19 | AAU01204 | Human p-dig protei |
| 36 | 228 | 4.4 | 1717 | 22 | AAU01204 | Novel human diago |
| 37 | 226.5 | 4.4 | 2048 | 22 | AAU01204 | Human polypeptide |
| 38 | 223.5 | 4.3 | 2400 | 22 | AAU01204 | Novel human diago |
| 39 | 220.5 | 4.3 | 1374 | 22 | AAU01204 | Human male enhance |
| 40 | 217.5 | 4.2 | 1975 | 22 | AAU01204 | Drosophila melanog |
| 41 | 217 | 4.2 | 1972 | 17 | AAU01204 | Smooth muscle myos |
| 42 | 217 | 4.2 | 2143 | 22 | AAU01204 | Novel human diago |
| 43 | 216.5 | 4.2 | 1372 | 19 | AAU01204 | Protein with Rho p |
| 44 | 212.5 | 4.1 | 1489 | 22 | AAU01204 | Drosophila melanog |
| 45 | 212.5 | 4.1 | 1988 | 22 | AAU01204 | Human polypeptide |

ALIGNMENTS

RESULT 1

AAE07164
ID AAE07164 standard; Protein; 1004 AA.

AAE07164;

06-NOV-2001 (first entry)

Human caspase recruitment domain-14 (CARD-14).

Human; caspase recruitment domain-14; CARD-14; chromosome 17;
nuclear factor-kappa B; NF-kB; cell growth; cell death; cancer; therapy;
autoimmune disorder; systemic lupus erythematosus; neurological disorder;
Alzheimer's disease; Parkinson's disease; inflammatory disorder; anaemia;
haematological disorder; myelodysplastic syndrome; myocardial infarction;
stroke; immune disorder; Crohn's disease; allergic rhinitis; infection;
cell signalling disorder; cytostatic; immunosuppressive; neurotropic;
neuroprotective; antiviral; antibacterial.

OS Homo sapiens.

| Key | Location/Qualifiers |
|---------------|---|
| Modified-site | 6..9 |
| Domain | /note= "CAMP- and cGMP-dependent protein kinase phosphorylation site" |
| Modified-site | 10..116 |
| Modified-site | /label= CARD_domain |
| Modified-site | 12..15 |
| Modified-site | /note= "Casein kinase II phosphorylation site" |
| Modified-site | 18..21 |
| Modified-site | /note= "Casein kinase II phosphorylation site" |
| Modified-site | 25..27 |
| Modified-site | /note= "Protein kinase C phosphorylation site" |
| Modified-site | 60..62 |

FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 91..93
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 114..117
 FT Modified-site /note= "N-glycosylation site"
 FT 117..122
 FT Modified-site /note= "N-myristoylation site"
 FT 121..123
 FT Domain /note= "Protein kinase C phosphorylation site"
 FT 126..420
 FT /label= Coiled_Coil_domain
 FT 130..135
 FT Modified-site /note= "N-myristoylation site"
 FT 134..137
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 161..166
 FT Modified-site /note= "N-myristoylation site"
 FT 165..168
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 220..227
 FT Modified-site /note= "Tyrosine kinase phosphorylation site"
 FT 221..224
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 239..325
 FT Domain /label= k-Box_domain
 FT 240..243
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 250..252
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 253..256
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 259..262
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 280..283
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 290..293
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 297..300
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 307..309
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 307..310
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 359..365
 FT Modified-site /note= "Tyrosine kinase phosphorylation site"
 FT 366..368
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 366..369
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 378..381
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 384..386
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 385..406
 FT Region /note= "Leucine zipper pattern"
 FT 449..452
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 463..466
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 463..465
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 470..472
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 501..504
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 511..516
 FT Modified-site /note= "N-myristoylation site"
 FT 568..660
 FT Domain /label= PDZ_domain
 FT 587..592
 FT Modified-site /note= "N-myristoylation site"
 FT 589..592
 FT Modified-site /note= "N-glycosylation site"

FT Modified-site 602..605
 FT /note= "Casein kinase II phosphorylation site"
 FT 634..637
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 653..655
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 674..677
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 676..745
 FT Domain /label= SH3_domain
 FT 714..719
 FT Modified-site /note= "N-myristoylation site"
 FT 725..727
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 725..728
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 733..738
 FT Modified-site /note= "N-myristoylation site"
 FT 737..740
 FT Modified-site /note= "N-glycosylation site"
 FT 759..761
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 760..763
 FT Modified-site /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
 FT 785..793
 FT Peptide /note= "Peroxisomal targeting signal"
 FT 796..799
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 800..805
 FT Modified-site /note= "N-myristoylation site"
 FT 826..1004
 FT Domain /label= Guanylate_kinase_domain
 FT 842..844
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 860..863
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 888..870
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 870..872
 FT Region /note= "RGD cell attachment sequence"
 FT 893..896
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 926..929
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 941..949
 FT Peptide /note= "Peroxisomal targeting signal"
 FT 944..947
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 976..979
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 980..985
 FT Modified-site /note= "N-myristoylation site"
 FT 1002..1004
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT
 FT WO200159065-A2.
 XX
 PN 16-AUG-2001.
 XX
 PD 22-JAN-2001; 2001WO-US02087.
 XX
 PF 09-FEB-2000; 2000US-0181159.
 XX
 PR (MILL-) MILLENNIUM PHARM INC.
 XX
 PA Bertin J;
 XX
 PI WPI; 2001-497073/54.
 XX
 DR N-PSDB; AAD13447.
 XX
 PT An isolated caspase recruitment domain polypeptide useful for regulating growth and cell death and useful for the treatment of cancer

[illegible]

CC systemic lupus erythematosus), neurological disorders e.g., Alzheimer's
CC and Parkinson's disease, inflammatory disorders, haematological disorders
CC (e.g., anaemia, myelodysplastic syndromes), myocardial infarctions,
CC strokes, immune disorders (e.g., Crohn's disease, allergic rhinitis),
CC cell signalling disorders and certain viral and bacterial infections.
XX
SQ Sequence 1139 AA;

Query Match 97.8%; Score 5034; DB 22; Length 1139;
Best Local Similarity 95.1%; Pred. No. 0;
Matches 995; Conservative 0; Mismatches 1; Indels 50; Gaps 3;

QY 1 MGELCRDSALTALDETLWEMMESHRIIVRCICPSRLTPPYLRQAKVLCQLEDEEVLSHS 60
DB 1 mgelcrrdsaltaldeetlwmemshrhriivrcicpsrltpyrlrqakvlcqledeevlshs 60
QY 61 PRITNSAMRAGHLDDLKTRGKNGAIAFLSKFHPNDVYTLVTGLQPDVDFSNFS---- 116
DB 61 pritnsamraghlldllktrgkngaiafleskfhpndvtytlvtglqpdvdfsnfsgess 120
QY 117 -----GLMETSKLTCAGAGTSGISLQELNKGKQEVLLRRCCOQLOE 158
DB 117 -----glmetskltecagagtsgislqelngkqevllrrccooloe 158
QY 121 dfdglagtsnrlrlvtcpqlmetsklteclagatgsiqeelngkqevllrrccqqlqe 180
DB 121 dfdglagtsnrlrlvtcpqlmetsklteclagatgsiqeelngkqevllrrccqqlqe 180
QY 159 HLGIAETRAEGLHQLEADHSMRKREVSAPHFVLRKDEMLSLSHYSNALQEKELAASR 218
DB 159 hlgiaetraeaglhlqleadhsmrkrevsahfhevlrkdemlsishysnalqekelaasr 240
QY 219 CRSLOEELYLKQELQANMVSCLELQELQSLRTASDQESGDEELNRLKEENKLSRLT 278
DB 241 crslqeelylkqelqanmvsscleeqlqslrtasdqesgdeelnrlkeeneklsrlt 300
QY 279 FSLAEKDILEOSDEARGSRQELVERIHSRERAVAAERQRE----- 320
DB 301 fslaekdileqsldeargsrqelverihsreravaaerqeqarpesllsftvvhshv 360
QY 321 QYWEKEQTLQFOKSKMACOLYREKYNALQAVCELOKEDQAYSARDSAQREISQSLV 380
DB 361 qyweekeqtlqfokskmacolyrekynalqavcelokerdqaysardsaqreisqslv 420
QY 381 EKDSLIRQVRELTDQVCELTQLROLQAEPPGVLLQKQARTPECPKQRLVRMHAIQPR 440
DB 421 ekdsirqvreltdqvceltqlrolqaeppgvllqkqartpecpkqrlvrmhaiqpr 480
QY 441 DSDCSLVSTFESQLLSDLSATSSRELVDTSFRSSPAPPQQSLYKRVAFDFGEEPWSFS 500
DB 481 ddsdcslvstfesqlldslsatssrelvdsfrssppappqsglykrrvaedfgeepwsfs 540
QY 501 SCLEIPEGDPGALPGAGDPHLDYELLDADLPQLESSLQPVSPGRLDYSE----- 552
DB 541 scliepegdpgalpgagdpghldyelladtadlpqlesslqpvspgrldvsesaqagrlp 600
QY 553 --SGVLMRRPARRILSQVTMLAFQGDALLBOISVIGNLTGIFTHRVTPGSAADOMALR 610
DB 601 acsgvlmrrparrilsqvtmlafqgdalleqisvignltgifihrvtvpgsaadqmalr 660
QY 611 PGTQIVMVDYEPASEPLFKAVLEDITLLEAVGLLRVRVDFGFCCLSVKVNVDGYKRLQDLEA 670
DB 661 pgtqivmvdyeapelfkavleditlleavglrrvdfgfcclsvkvntdgykrlldlea 720
QY 671 KVATSGDSFYTRVNLAMEGRAGKELQVHCNEVLHVDTMTFGCCGCHWAHRVNSYTMKDTA 730
DB 721 kvatsgdsfytrvnlemeagrakelqvhcnevhlvdtmtfmgcgchwhahrvnstymkda 780
QY 731 AHGTPNYSRAQQOILIALIQDMTOCTVTRKPPSGGPKLVRIVSMKAKASPLRLSFDR 790
DB 781 ahgtipnysraqqoiliqliqdmtoctvtrkpsgggpklvriivsmkakasplrlsldr 840
QY 791 GOLDPSPRMGSGSTCFWAFESCITLVPYTLVPHRPARPRPVLLVPRVAGKILSEKCLLQOG 850
DB 841 gqldpsrmegsgstcfwafescitlvpvtlvrphrparprpvllvpravgkilseklcllqg 900

QY 851 FKKCLAEYLSQBEYEWNSORGDIQEGEVSGRCWVTRHAVESLMEKNTHALLDVQSDSV 910
DB 901 fkkclaeylsqbeeyawsrgdiqegevsgrcwvtrhaveslmeknthalldvqsdsv 960
QY 911 CTLHRMDIFPIVIVHSVNEKMAKLLKKGILQRLGTSEEQLLEAARQEGDLDRAPCLYSSL 970
DB 961 ctlhrmdifpivihsvnekmakllkkgilqrlgtseeqlleaaarqegdlrapclyssl 1020
QY 971 APDGSWSDLGLLSCVRQAIADQKKV 996
DB 1021 apdgswdlgllscvraqiadeqkv 1046
RESULT 3
AAU01207
ID AAU01207 standard; Protein; 1147 AA.
XX AAU01207;
AC AAU01207;
XX 12-SEP-2001 (first entry)
XX Human caspase recruitment domain, CARD-11 polypeptide.
XX Human; caspase recruitment domain; CARD-11; Bcl-10; NF-kappaB;
KW apoptosis; hyperproliferative disorder; autoimmune; neurological;
KW inflammatory disorder; viral infection; stress-related response.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 6..112
FT Modified-site /note= "CARD domain"
FT Modified-site /note= "Protein kinase C phosphorylation site"
FT Modified-site /note= "Casein kinase II phosphorylation site"
FT Modified-site /note= "Casein kinase C phosphorylation site"
FT Modified-site /note= "Protein kinase C phosphorylation site"
FT Modified-site /note= "Casein kinase II phosphorylation site"
FT Modified-site /note= "Protein kinase C phosphorylation site"
FT Modified-site /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Domain 130..431
FT Modified-site /note= "Coiled coil domain"
FT Modified-site /note= "Casein kinase II phosphorylation site"
FT Modified-site /note= "Casein kinase II phosphorylation site"
FT Modified-site /note= "Tyrosine kinase phosphorylation site"
FT Modified-site /note= "Casein kinase II phosphorylation site"
FT Modified-site /note= "Tyrosine kinase phosphorylation site"
FT Modified-site /note= "N-glycosylation site"
FT Modified-site /note= "Protein kinase C phosphorylation site"
FT Modified-site /note= "Amidation site"
FT Modified-site /note= "Casein kinase II phosphorylation site"
FT Modified-site /note= "Protein kinase C phosphorylation site"
FT Modified-site /note= "Casein kinase II phosphorylation site"
FT Modified-site /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 459..461

Db 71 dilhtkgqgyvvfleslefyypelyklygtkeptrfstivveegheglthflmnevik 130
QY 135 LOEELNQERQGVKQVLLRRQQQLQ-EHLGLAETRAEGLHQAEDHSHRMKREVSAPHEVLR 193
Db 131 lqqmkakdlqrcellarlrqledekqmtltrve-lltfqeryykmkeersyndelvk 189
QY 194 LKDEMLSLSLHYSNALQERELAASRCRSLQEBELLYLLKQBLQANMVSSCELELEQOSLRT 253
Db 190 vkddnynlamryaqiseeknmavmsrdqlgleldqikhrink--meeckle-rnqslk 246
QY 254 ASDOES--GDEELNRUKERNEKRLSTFSL-----AEKDILEOSLDEAR 295
Db 247 kndienrpkeqglelerenemlktknqqlsqiagkrsldpsdkailidilehdkreal 306
QY 296 GSRQELVERIHSRLRAVAAREQREYWEKEQTLQFOKSMACOLYREKYNALQAQVC 355
Db 307 edrqelvnriynlqeeaqeeldkyleekedlelkctlgkdcemykhrmtvmlgle 366
QY 356 ELOKERDQAYSARDSAQRISQSLVEKDSLRQVPELTQO-----VCELR 401
Db 367 evererdqghardeaqtgysqcliekkyrkqireleekndemremvrrreacivnles 426
QY 402 QLROLOAE-----PGVLVKQ-----EARTREPCPREK----- 428
Db 427 klrrlskdsnnldqslprnlpvtiisqdfgdasprtnqgeadstseespedskyflpy 486
QY 429 ---QRLVRMHAI-CPRDDSDCSLVSTESQLL-----SDLATS---SRELVDSPFRSSP 476
Db 487 hppqrrmnikqlqrakqspislrktsdfqagheegtdaspsscgslpitnsffkmpq 546
QY 477 -----APPQOSLYKRVAFEDFEEFSPSCLEIEPGDGPALPGAKA-CDPHL 523
Db 547 prsrssltaeppgndsvivrykedaphr-----stve-edndsggfdaldlddshe 600
QY 524 DY-----ELLDTADLPQLESSLPQVS---PGRLDVSESGVLMRRRPAR 564
Db 601 rysfggssihssssshqsegldayleqvnlmfrkfslerfpvstsvghvrgpps-- 658
QY 565 ILSQVTMLAFQGDALIEQISVIGNGTGFIHRVTPGSAADOMALRPGTOIVMVDYEASE 624
Db 659 ----vqhttdngsltsqitllggnargsvfshvkgpslaekagregbqllllegcirg 714
QY 625 PLFKAVLEDTLEAVGLLRVDFCCLSVKVNTDCYKRLQDLEAKVATSGDSFVIRVN 684
Db 715 ergsvpldtcteeahwtlqrcspgvtlhykvnhegyrkivkmedglitsgdsfyirln 774
QY 685 LAMEGRAGK-ELQVHCNEVLHVTDTFMQCGCWHHRVNSYTMKDTAAHGTIPNYSRAQ 743
Db 775 lnissqldactmslkcdvvhvrdtnyqdrhewpcarvpdftdhdldm-gtipsysraq 833
QY 744 QLTALQDWTQO-----CVTRKP-----SSGGPQ-----KL 771
Db 834 lllvklqrlmhrgrsreevdgthlralrntlqpealstsdprvsrlsraflgqll 893
QY 772 RIVSMDKAK-----ASPL-----RLSEF-----RGOLDPSRMEGSGSTCF 805
Db 894 qfvsrsenkykrnmnsnervitssgplslarslslatklkteqelopeselgkn--- 950
QY 806 WAESCTLPVPTLVWPHRRPARPVLVRAVGKILSEKILCLLQO---FKKCLAEYLSQE 862
Db 951 -----lslpyslvrafycerrrpvlftptvltqvlrllnsggameficksdivtrd 1005
QY 863 EYBWSQORGDI--QSGEVSGGRCWTRHVAVESLMKBNTHALLDOLDVSDVCTLHRMDIPP 920
Db 1006 efllrqktetliysreknpnafec-apanieavaaknhclleaglgctrdliknlyp 1064
QY 921 IVIHVSVNEKMAKLGKGLQRLCTSEQLLEAARQEGDLDRAPCLYSSLPADGWSLDLG 980
Db 1065 ivlfrvceknikrfrklprpet-eeeflrvcrlkekealealpclyatvepdmwgsvee 1123
QY 981 LLSVCVRQAIADQKVVVWTEQ 1001
Db 1124 llrvvdkdgeqrktiwdde 1144

RESULT 4
AAU01206
ID AAU01206 standard; Protein; 1032 AA.
XX AAU01206;
AC AAU01206;
XX 12-SEP-2001 (first entry)
XX Human caspase recruitment domain, CARD-10 polypeptide.
XX Human; caspase recruitment domain; CARD-10; Bcl-10; NF-kappaB;
KW apoptosis; hyperproliferative disorder; autoimmune; neurological;
KW inflammatory disorder; viral infection; stress-related response.
XX Homo sapiens.
XX Key Location/Qualifiers
FH Modified-site 15..20 /note= "N-myristoylation site"
FT Modified-site 18..21 /note= "Casein kinase II phosphorylation site"
FT Domain 23..123 /note= "CARD domain"
FT Modified-site 68..70 /note= "Protein kinase C phosphorylation site"
FT Modified-site 76..79 /note= "N-glycosylation site"
FT Modified-site 78..80 /note= "Protein kinase C phosphorylation site"
FT Modified-site 88..91 /note= "Amidation site"
FT Modified-site 112..115 /note= "Casein kinase II phosphorylation site"
FT Modified-site 113..118 /note= "N-myristoylation site"
FT Domain 147..457 /note= "Coiled coil domain"
FT Modified-site 201..207 /note= "Tyrosine kinase phosphorylation site"
FT Region 230..251 /note= "Leucine zipper homology region"
FT Modified-site 242..245 /note= "Casein kinase II phosphorylation site"
FT Modified-site 293..295 /note= "Protein kinase C phosphorylation site"
FT Modified-site 293..296 /note= "Casein kinase II phosphorylation site"
FT Modified-site 309..314 /note= "N-myristoylation site"
FT Modified-site 313..315 /note= "Protein kinase C phosphorylation site"
FT Modified-site 331..334 /note= "Casein kinase II phosphorylation site"
FT Domain 366..398 /note= "Tropomyosin domain"
FT Modified-site 412..415 /note= "Casein kinase II phosphorylation site"
FT Region 426..447 /note= "Leucine zipper homology region"
FT Modified-site 438..441 /note= "Casein kinase II phosphorylation site"
FT Domain 457..1032 /note= "MAGUK domain"
FT Modified-site 472..475 /note= "N-glycosylation site"
FT Modified-site 478..481 /note= "Casein kinase II phosphorylation site"
FT Modified-site 487..492 /note= "N-myristoylation site"
FT Modified-site 510..513 /note= "Casein kinase II phosphorylation site"

QY 770 LVRIYSMD-----KAKAPRLSLFRDQLDPSRMGSSSTCFWAEBSCLTLVPTLVWPH 822
 Db 505 nlkkraldqlrlvrpkpgpagagdpdqllep-----c--aeperslrpyslvrl 554
 QY 823 RPAPRPVLLVPRANGKILSEKLCILQG----FKKCLABYLQSEYEAWSQRDIIQGE 878
 Db 555 lvsalrpvllpeclaprlrnllldpsrlldfvcpaeslsgeelcspasgpaqka 614
 QY 879 VSGRCWVTRHAVESLMEKNTHALLDVQLDSVCTLHRMDIFPIVTHVSVNEKMAKLKKG 938
 Db 615 tpglgsri-raiqesvgkk--hcllelgargvrelvqneipvlhvvecknrevrql 671
 QY 939 LQRLGTSEQLLEAARQEGDLDRAPCLYSSLPAGDGLDGLLSCVROAIADEQKVVW 998
 Db 672 lrgpwrdsellrqrcgseqvlgwlpeswvqpahewghaeelakvvrllqeqarlw 731
 QY 999 TE 1000
 Db 732 ve 733

RESULT 6
 AAU01204
 ID AAU01204 standard; Protein; 536 AA.
 AC AAU01204;
 XX
 XX
 DT 12-SEP-2001 (first entry)
 XX
 DE Rat caspase recruitment domain, CARD-9 polypeptide.
 XX
 KW Rat; caspase recruitment domain; CARD-9; Bcl-10; NF-kappaB;
 KW apoptosis; hyperproliferative disorder; autoimmune; neurological;
 KW inflammatory disorder; viral infection; stress-related response.
 XX
 OS Rattus sp.
 XX

Key Location/Qualifiers
 FH Modified-site 2..5
 FT /note= "Casein kinase II phosphorylation site"
 FT Domain 7..98
 FT /note= "CARD domain"
 FT Modified-site 12..15
 FT /note= "Casein kinase II phosphorylation site"
 FT Modified-site 16..18
 FT /note= "Protein kinase C phosphorylation site"
 FT Modified-site 23..26
 FT /note= "Casein kinase II phosphorylation site"
 FT Modified-site 92..95
 FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
 FT Modified-site 95..97
 FT /note= "Protein kinase C phosphorylation site"
 FT Modified-site 95..98
 FT /note= "Casein kinase II phosphorylation site"
 FT Modified-site 138..140
 FT /note= "Protein kinase C phosphorylation site"
 FT Modified-site 138..141
 FT /note= "Casein kinase II phosphorylation site"
 FT Domain 140..416
 FT /note= "Coiled coil domain"
 FT Modified-site 171..174
 FT /note= "Casein kinase II phosphorylation site"
 FT Modified-site 176..183
 FT /note= "Tyrosine kinase phosphorylation site"
 FT Region 197..213
 FT /note= "Indole-3-glycerol phosphate synthase homology region"
 FT Modified-site 228..231
 FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
 FT Modified-site 231..233

FT Modified-site
 FT /note= "Protein kinase C phosphorylation site"
 FT 267..270
 FT /note= "Casein kinase II phosphorylation site"
 FT 285..338
 FT Region
 FT /note= "Cysteine rich repeat homology region"
 FT 303..305
 FT Modified-site
 FT /note= "Protein kinase C phosphorylation site"
 FT 362..364
 FT Modified-site
 FT /note= "Protein kinase C phosphorylation site"
 FT 362..365
 FT Modified-site
 FT /note= "Casein kinase II phosphorylation site"
 FT 374..377
 FT Modified-site
 FT /note= "Casein kinase II phosphorylation site"
 FT 425..428
 FT Modified-site
 FT /note= "Casein kinase II phosphorylation site"
 FT 431..433
 FT Modified-site
 FT /note= "Protein kinase C phosphorylation site"
 FT 451..453
 FT Modified-site
 FT /note= "Protein kinase C phosphorylation site"
 FT 483..486
 FT Modified-site
 FT /note= "Casein kinase II phosphorylation site"
 FT 514..516
 FT Modified-site
 FT /note= "Protein kinase C phosphorylation site"
 FT 523..528
 FT Modified-site
 FT /note= "N-myristoylation site"
 FT 524..527
 FT Modified-site
 FT /note= "N-glycosylation site"
 FT 526..529
 FT Modified-site
 FT /note= "Casein kinase II phosphorylation site"
 FT 531..534
 FT Modified-site
 FT /note= "Casein kinase II phosphorylation site"
 FT
 XX WO200140468-A2.
 PN
 XX
 PD 07-JUN-2001.
 XX
 PF 01-DEC-2000; 2000WO-US32716.
 XX
 XX 03-DEC-1999; 99US-0168780.
 PR 18-FEB-2000; 2000US-0507533.
 PR 25-FEB-2000; 2000US-0513904.
 PR 10-OCT-2000; 2000US-0685791.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Bertin J;
 XX
 DR WPI; 2001-367809/38.
 DR N-PSDB; AAS05386.
 XX
 PT Novel caspase recruitment domain (CARD) proteins, CARD-9, CARD-10,
 PT CARD-11, useful as targets for therapy, as immunogens, and in screening
 PT and detection assays -
 XX
 PS Claim 9; Fig 1A-1B; 145pp; English.
 XX
 CC The present sequence represents novel rat caspase recruitment
 CC domain, CARD-9. The polynucleotide encoding this sequence was
 CC isolated from a rat neuronal cDNA library. Also described are novel
 CC human sequences for CARD-9, CARD-10 and CARD-11 (AAU01205-AAU01207).
 CC CARD-9, CARD-10 and CARD-11 interact with Bcl-10 which is thought
 CC to activate nuclear factor (NF)-kappaB and apoptosis. The sequences
 CC of the invention can be used for treating a disorder associated
 CC with abnormal levels of apoptosis by modulating the expression
 CC or activity of CARD-9, CARD-10, or CARD-11. They can be used for
 CC the treatment of hyperproliferative disorders (e.g. cancer),
 CC autoimmune disorders (e.g. systemic lupus erythematosus), neurological
 CC disorders (e.g. Alzheimer's disease), inflammatory disorders (e.g.
 CC Crohn's disease), and viral infection (e.g. HIV). The CARD polypeptide,
 CC polynucleotide and an antibody which selectively binds to CARD can be
 CC used in screening and detection assays (e.g. chromosomal mapping, tissue
 CC typing), predictive medicine (prognostic assays, monitoring clinical
 CC trials, and therapy (treatment and prophylaxis). The CARD polypeptide may

CC be used to screen for drugs that bind to and/or modulate it. CARD
CC sequences are potential targets for regulating inflammation, cancer,
CC NF-kappaB signalling, stress-related response and apoptosis in human
CC disease. A host cell containing a polynucleotide encoding CARD can be
CC used to create transgenic animals.

XX Sequence 536 AA;

Query Match 10.0%; Score 512.5; DB 22; Length 536;
Best Local Similarity 28.8%; Pred. No. 2.4e-33;
Matches 162; Conservative 94; Mismatches 215; Indels 91; Gaps 15;
QY 15 DEETLWMMESHRRIVRCICPSRLPYLRQAKVLCQDEEEVLHSPRLNSAMRAGHLL 74
Db 6 nddecwsalesfrvklisvidpsritpylrqckvlnpddeeqvlsdpnlvirkrkvgl 65
QY 75 DLLTRCKNGAIAFLSKLPHNDVTLVTGLQPD-----VDFSNFSGL-----ME 120
Db 66 dlilqrthggyvafleslelypplyrkyvtgkparvfsmildasgesgitqlmtvmk 125
QY 121 TSKLTELAGAIGSLQELNQEKGQEVLLRRCQQQLQEHGLGAEAEGLRQLLEADHSRM 180
Db 126 lqkvqdltais-kdfikelirvksdlirkhqervq-----rll 164
QY 181 KREVSAPHFVBLKDEMLSLHYSNALQEKELASRCSLQELLYLLKQELQRANMWS 240
Db 165 keecelssaelkrckdenydlamrlahlseekgaalnrldlqlevdqlrshlmkae--d 222
QY 241 SCLELEQEQ-SLRASDOESGDEELNRLKEENE---KLRSLLTFLSAEK-----DI 286
Db 223 dckverkhtklrhameqrpsqellwdlqgdrllqarvdelevsqegklhrnsyiqv 282
QY 287 LEQSLDEARGSRQELVERIHSRLRERAVAAERQEQWEEKETLLQFQKSKMACQLYREK 346
Db 283 leedwralqehqegastifslrkdrlqaealtrcmeeekemfclqlairkdakmykdr 342
QY 347 VNALQAVQVCLQEQVQAYSARSAQREISQSLVSKDLSLRQVFEITDQVCELTQLRQL 406
Db 343 ieailqgmeevslerdqamsteelhaqcagsfgdklkrkvreidekadqlqlqifqt 402
QY 407 QAEPPGVVKQEAETREPCPREKORLV---RWAHAICPRDDSDCSLSVSTESQL-LSLSLAT 462
Db 403 esr---llaaegrkq---qqlmllsldesprnsqelslpqdleedaglsdkgvl 456
QY 463 SSRE-----LVDSFRSSSPAPPQSQSLYKRVAEDFG-----EEPWS 498
Db 457 adrespeqpfvlnkkhlsqthdtpvsssepekeerrlkesfenyrrkralrkmgnsr 516
QY 499 FSSCLEIPEGDPGALPCAKAGD 520
Db 517 qg-----egdhgnttgdsdntd 532

RESULT 7
AAU01205
ID AAU01205 standard; Protein; 536 AA.
XX AC AAU01205;
XX DT 12-SEP-2001 (first entry)
XX DE Human caspase recruitment domain, CARD-9 polypeptide.
XX KW Human; caspase recruitment domain; CARD-9; Bcl-10; NF-kappaB;
KW apoptosis; hyperproliferative disorder; autoimmune; neurological;
KW inflammatory disorder; viral infection; stress-related response.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Modified-site 2..5
FT /note= "Casein kinase II phosphorylation site"

FT Domain 7..98
FT /note= "CARD domain"
FT Modified-site 23..26
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 92..95
FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 95..97
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 95..98
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 138..140
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 138..141
FT /note= "Casein kinase II phosphorylation site"
FT Domain 140..416
FT /note= "Coiled coil domain"
FT Modified-site 176..183
FT /note= "Tyrosine kinase phosphorylation site"
FT Region 137..213
FT /note= "Indole-3-glycerol phosphate synthase homology region"
FT Modified-site 228..231
FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 231..233
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 267..270
FT /note= "Casein kinase II phosphorylation site"
FT Region 285..338
FT /note= "Cysteine rich repeat homology region"
FT Modified-site 303..305
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 333..336
FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 363..366
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 425..428
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 431..433
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 450..452
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 453..458
FT /note= "N-myristoylation site"
FT Modified-site 460..462
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 481..486
FT /note= "N-myristoylation site"
FT Modified-site 483..486
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 524..527
FT /note= "N-glycosylation site"
FT Modified-site 526..529
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 527..532
FT /note= "N-myristoylation site"
FT Modified-site 531..534
FT /note= "Casein kinase II phosphorylation site"
XX WO200140468-A2.
XX 07-JUN-2001.
XX 01-DEC-2000; 2000WO-US32716.
XX 03-DEC-1999; 99US-0168780.
XX 18-FEB-2000; 2000US-0507533.
XX 25-FEB-2000; 2000US-0513904.
XX 10-OCT-2000; 2000US-0685791.
XX (MILL-) MILLENNIUM PHARM INC.


```
Db 263 llnkgaasplepkeleelrdkn-----estlmrlhetlkqcqdlkteksqmdrk 312
QY 64 TNSAMRAGHLDDLKTRGKNGAIATF-LESUKFHNPDVYTLVTLGQDPDVFNSFSGLMETS 122
Db 313 inqlse-----engdisfklrefashlqlqldalnelteehskatqewlekqa 360
QY 123 KLTECLAGATGSLQELNEQKQGVLLRRCQQLQEHGLAETRAEGLHQLEADHSMRKR 182
Db 361 qleleisaa---lqdkkcllee-kneillqgklsqleehls-----qlqdnppqekg 406
QY 183 EVSAHFHEVLRLKDEMLSL-----LHYSNALQEKELAAASRCRSLOEELYL-- 228
Db 407 evlgdvqlqetlkqeaatlaanntqlqarvemleterggqekllaerghfeekqgls 466
QY 229 -----LKOELQRANNVSSCELELOESLRT-----ASDQESGDEELNRLK 268
Db 467 lltlqssisnlsgakeeqasqahgaritacqvasltelttnatigqq-dqelaglk 525
QY 269 EE-----NEKLSLTFSLAEKDILEQSLDEARGSRQELVERIH 306
Db 526 qgakekqaglaqltqdgqasqgrrhqveqlssllkqk---eqqlkev-aekqeatrqdh 581
QY 307 SLRERAVAAERQEQWEEKQOTLLOFQSKMACQLYREKVNALQAVQVCLEQKEROQAYS 366
Db 582 a-qqlataae-ereaslrerdaalkle-----alekekaaklei---lqqqlqvane 629
QY 367 ARDSAREISQSIVKXDSLRQVFEL-----TDQVCELRTQLROLQAEPPG 412
Db 630 ardsaqsvtqagreqaelskrveeqacvetarqeqheaqvaqaeleqlrseq----- 684
QY 413 VLQOEARTREPCPREKORLYRMAHCPRDDSDCSLVSSTESQLLSLDSATSSRELVDSEF 472
Db 685 ---gkatekervaqekdql-----qeqqlakeslkvikgsleeeekr 723
QY 473 SSSPAPPSQSLKRVAAEDFGEPPWFSFSCLEIPEGDPGALPGAKAGADPHLDYELDLTAD 532
Db 724 raadaleeqrciselkae-----trslveqhkterkeleeragrkqlearllqge 776
QY 533 LPQLESSLPVSPGRLDVSESGVLMRRRPARILTSQVITMLAFQGDALLQISVIGNLGT 592
Db 777 ahaete-----vlrrelaamaagt-----aesceql----- 806
QY 593 IFTHRTVPGSAADQMALPCTQIVMVDYEAEPLEFRAVLEDTTLEEAAGVLLRRVDGFCC 652
Db 807 -----vkeavaawrdye-----dsqeeaqygamfgeq-----lm 836
QY 653 SVKVTNMGYKRLQLDLQLEAKVATSGDSFYIRVNLAMEGRAKGELOVHCNEVLHVTDTMFQ 712
Db 837 tlkeecekarqelqeaekva--gieshselqisrqgnklael--han----- 880
QY 713 CGCWHHRVNSYTMKDTAAHGTIPNYSRAQQOLIALIQTMTQCTVTRKPSGGPQKLV- 771
Db 881 ----laralqvgqekvraqdaddistlqekmaatskevarletlvrk--ageeqetas 934
QY 772 RIVSMDKAKASPLRSLPDRGOLDPSPRME---GSSTCFWAESCLTLVPYTLVWPHRRPARPR 828
Db 935 relvkepara-----gdrpewleeeqggrqfc-----stqa 965
QY 829 PVLIVPRAVKILSEKLCILQGGKCLAEYLSOEYEWAMSORGDIQEGEVSGRCWTR 888
Db 966 alqamereaeqmgne---lerlraalmesggqgeergqgevariltqerg-----r 1015
QY 889 HAVESLMEKNYTHALLVDLOSVCVTLHRMDIFPI--VLHVSVNEKMAK-----KLK----- 936
Db 1016 aqadlalekaaraaelmrlnalnegrvfatigqaelahaltekgkdqelakirgleaa 1075
QY 937 --KGLQRLGTSEEOLLEAARQEE 957
Db 1076 qikeleelrqtvtqkqleqlakke 1098
```

RESULT 12

```
AAG65799
ID AAG65799 standard; protein; 2101 AA.
XX
AC AAG65799;
XX
DT 30-JAN-2002 (first entry)
XX
DE Nuclear mitotic apparatus protein (NUMA) sequence.
XX
KW NuMA; Nuclear Mitotic Apparatus protein; malignant; cell growth; human;
KW tumor progression; structural protein.
XX
OS Homo sapiens.
XX
PN US6287790-B1.
XX
PD 11-SEP-2001.
XX
PF 30-NOV-1999; 99US-0452294.
XX
PR 30-NOV-1998; 98US-110420P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Lelievre S, Bissell M;
XX
DR WPI; 2001-610603/70.
XX
CC Distinguishing growth arrested malignant cells from proliferating
CC non-malignant cells, comprises staining the nuclear mitotic apparatus
CC protein in intact mammalian cells and determining the 3-dimensional
CC pattern of the labeled protein.
XX
PS Disclosure; Fig 15; 38pp; English.
XX
CC The invention relates to methods of distinguishing malignant or growth
CC arrested malignant cells from proliferating non-malignant cells by
CC staining specifically Nuclear Mitotic Apparatus (NUMA) protein in a
CC sample of intact mammalian cells and imaging the cells to determine the
CC three-dimensional pattern of labeled NUMA protein within nuclei of the
CC cells. The method is useful for distinguishing malignant or growth
CC arrested malignant cells from proliferating non-malignant cells.
CC Localization of structural proteins such as NUMA may be used to identify
CC tumors cells and different stages in the tumor progression and
CC differentiation processes. NUMA proteins may be used to transport
CC functional proteins or drugs to specific locations inside the cells.
CC The present sequence represents the human NUMA protein (GenBank Accession
CC Number Z11384).
XX
SQ Sequence 2101 AA;
```

Query Match 5.1%; Score 262; DB 22; Length 2101;
Best Local Similarity 19.9%; Pred. No. 1.le-11;
Matches 208; Conservative 163; Mismatches 376; Indels 296; Gaps 39;

```
QY 4 LCRRDALTALDEETLWEMMESHRHRTVRCICPSRLTPYLRQAKVLCQDLDEEVLHSPRL 63
Db 263 llnkgaasplepkeleelrdkn-----estlmrlhetlkqcqdlkteksqmdrk 312
QY 64 TNSAMRAGHLDDLKTRGKNGAIATF-LESUKFHNPDVYTLVTLGQDPDVFNSFSGLMETS 122
Db 313 inqlse-----engdisfklrefashlqlqldalnelteehskatqewlekqa 360
QY 123 KLTECLAGATGSLQELNEQKQGVLLRRCQQLQEHGLAETRAEGLHQLEADHSMRKR 182
Db 361 qleleisaa---lqdkkcllee-kneillqgklsqleehls-----qlqdnppqekg 406
QY 183 EVSAHFHEVLRLKDEMLSL-----LHYSNALQEKELAAASRCRSLOEELYL-- 228
Db 407 evlgdvqlqetlkqeaatlaanntqlqarvemleterggqekllaerghfeekqgls 466
QY 229 -----LKOELQRANNVSSCELELOESLRT-----ASDQESGDEELNRLK 268
```

```
Db 467 litdlqssisnlsqakeeegagaharltaqvasiltseittlnatiqq-dqelaglk 525
QY 269 EE-----NEKLSLTFSLAEKIDLEQSLDEARGSRQELVERIH 306
Db 526 qqakekqaqlatlgqgeasgglrhqveqlsslskq---eqqlkev-aekqeatrqdh 581
QY 307 SLRERAVAAERQEWEEKOTLLQFOKSKMACOLYREKVNALQAVCELOKQERDOAYS 366
Db 582 a-qqlataae-ereaslterdaalkle-----alekekaaklei---lqqqlqvane 629
QY 367 ARDSAQREISQSLVEKDSLRROVFEL-----TDQVCELRTQLROLQAEPGP 412
Db 630 ardsqtsvtgqrekaelskrveelqacvetarqdehaqgaevaelqlrseq----- 684
QY 413 VLKQARTREPCPREKORLRVMAHCPRDDSDCLSVSPFESOLLSDLSATSSRELVDSPR 472
Db 685 ---qakekervaqekdql-----geqlqalkeslvtkgsleeekr 723
QY 473 SSSPAPPQQSLYKRVADFGEEPWFSFSCLEIPEGDPGALPKAKAGADPHLDYELDTAD 532
Db 724 raadaleeqgrciselkae-----trslveqhkrrerkeleeeagrkglearllqige 776
QY 533 LPQLSSSQPVSPGRLDVSESVLMRRPARILSOVTMLAFQGDALLLEQISVIGNLTG 592
Db 777 abqaete-----virrelaeamaagt-----aeseceql----- 806
QY 593 IFIHRVTPGSAADQALRGCTQIVWVDVEASEPLFKAVLEDTTLEAVGLLRRVDGFCCL 652
Db 807 -----vkevaawrdgye-----dsqgeaqgymfgeq-----lm 836
QY 653 SYKVNTDGYKRLQLQLEAKVATSGDSFYIRVNLAWEGRAGKELQVHCNEVLHVTDTMFGQ 712
Db 837 tlkeeeekarqelqakekva--gleshelqisrqnklael--han----- 880
QY 713 CCQWIAHRVNSTYMDTAHGTIPNYSRAQQQLIALIQDWTQCTVTRKPPSSGGQPKLV- 771
Db 881 ----laralqvgqevraqkladdlstlqekmaatskevarletlvrk--agedqetas 934
QY 772 RIVSMDKAKAPLRLSDFRGQLDPSRME---GSSTCFWAEASCLTLVPTLVWPHRPARPR 828
Db 935 relvkepara-----gdrqpeveeqgrqfc-----stqa 965
QY 829 PVLVPRVAVGKYLSEKLCILLOGFKCLAEYLSQEEYEAWSQRGDIQGEVSGGRCWTR 888
Db 966 alqamereaeqmgne----lerlraalmesggqgeergqerevarltqerg-----r 1015
QY 889 HAVESLMENKTHALLDVQLDSVCTLHRMDIFPI--VIHVSVNEKMAK-----KLK----- 936
Db 1016 aqadialaekaaraelemrignalneqrvefatlqealahaltekegkdqelaklrgleaa 1075
QY 937 --KGLQLGTSPEQLLEAARQEE 957
Db 1076 qikeleelrtqvklqkeqlakke 1098

RESULT 13
AAU32041
ID AAU32041 standard; Protein; 2207 AA.
XX
AC AAU32041;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #2532.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
```

```
XX 25-OCT-2001.
PD
XX 16-APR-2001; 2001WO-US08656.
PF
XX 18-APR-2000; 2000US-0552929.
PR 26-JAN-2001; 2001US-0770160.
XX (HYSE-) HYSEQ INC.
PA
XX Tang YT, Liu C, Drmanac RT;
PI WPI; 2001-611725/70.
XX
DR Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
PS Claim 20; Page 547; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful
CC for determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.
XX
SQ Sequence 2207 AA;
```

```
Query Match 5.1%; Score 262; DB 22; Length 2207;
Best Local Similarity 19.9%; Pred. No. 1.1e-11;
Matches 208; Conservative 163; Mismatches 376; Indels 296; Gaps 39;

QY 4 LCRRDSALTALDDEETLWEMESHRRIVRCICPSRLTPYLROAKVLCQLDDEEVLHSPRL 63
Db 269 llinekqaasplepekeleirdxn-----esltmrhletikqcdlkteksqmdrk 318
QY 64 TNSAMRAGHLDDLKTRGNKAIAF--LESKLFHNPDPVYTLVTGLQPDVDFSNFGLMETS 122
Db 319 inqlse-----engdisfkrlrefashlqqlqldalnelteehskatqewlekqa 366
QY 123 KLTECLAGAIGSLQOEELNQEKGKEVLLRRCQQLQOEHLGLAETRAEGLHOLEADHSMRKR 182
Db 367 qiekelsaa---lqdkkcllee-kneillqgklslqeels-----qlgnppqekg 412
QY 183 EYSAHPFVRLRLKDEMLSLS-----LHYSNALQEKELAAASRCSLQDELYL-- 238
Db 413 evlgdvlqietikqeaatlaanntqlqarvemleterggqekallaerghfeekqqlss 472
QY 229 -----LKOELQRANMVSSCELEQEQSLRT-----ASDQSGSDEELNRLK 268
Db 473 litdlqssisnlsqakeeegagaharltaqvasiltseittlnatiqq-dqelaglk 531
QY 269 EE-----NEKLSLTFSLAEKIDLEQSLDEARGSRQELVERIH 306
Db 532 qqakekqaqlatlgqgeasgglrhqveqlsslskq---eqqlkev-aekqeatrqdh 587
QY 307 SLRERAVAAERQEWEEKOTLLQFOKSKMACOLYREKVNALQAVCELOKQERDOAYS 366
Db 588 a-qqlataae-ereaslterdaalkle-----alekekaaklei---lqqqlqvane 635
QY 367 ARDSAQREISQSLVEKDSLRROVFEL-----TDQVCELRTQLROLQAEPGP 412
```

Db 636 ardsaqsvtqgarekelskrveelqacvetarqeqheaqgvaeeleqlrseq----- 690
 QY 413 VLKQEARTEPCPCORLVRMHAIACPRDDSDCLSVSSTESQLLSDLSATPSSRELVDSPR 472
 Db 691 ---qkatekervaqekdq-----qeqalkeslvtkgslseekr 729
 QY 473 SSSPAPPSQSLKYRVAEDGEPWFSSCLEIPGDPGALPGAKAGDHPHLYELLDTAD 532
 Db 730 raadaaleeqrciselkae-----trslveqhkrrerkeleeragrkglearllqlge 782
 QY 533 LPQLESSLPVSGRLDVSFSGVLMRRRPARRILSQVTMLAFQGDALLESISVIGGNLTG 592
 Db 783 ahgaete-----vrlrelaeamaaght-----aeseceql----- 812
 QY 593 IFIHRVTPGSAADQMALPGQTQVMVDYEAASEPLFRKAVLEDDTTLEBAVGLLRVDFGFCCL 652
 Db 813 -----vkevaawrdgve-----dsqceaqygamfgeq-----lm 842
 QY 653 SVKVNVDGYRLLQDLLEAKVATSGDSFYIRVNLAMEGRAKGELOVHCNEVLHVTDMFQG 712
 Db 843 tlkeceekargelqeakekva--gieshselqisrqnklael--han----- 886
 QY 713 CGCWHHRVNSYTMKDTAAHGTTIPNYSAQQOOLIALIODMTQCTVTRKPSGGPOKLV- 771
 Db 887 ----laraiqqvqekevrakgladdlstlqekmaatskevarletivrk--ageqgetas 940
 QY 772 RIVSMOKAKASPLRSLFDRQQLDPSRME---GSSTCFWAEBSCLTLVPTYLWPHRPARPR 828
 Db 941 relvkepara-----gdrpewleeeqgrqfc-----stqa 971
 QY 829 PVLVPRVAKGILSEKLLCLQGFKKLAELYLSQEEFVANSQSGDIQEGEVSGRGWVTR 888
 Db 972 alqamereaeqmne-----lerlraalmesqgqgeqgqerevarltdqerg-----r 1021
 QY 889 HAVESLMKNTNTHALLDVLDSVCTLHRMDIFPI--VIHVSVNEMKMAK-----KLK----- 936
 Db 1022 aqadlalekaaraelemrlqalneqrvefatigaelahaitkekgdkqlaklrgleaa 1081
 QY 937 --KGLQRLGTSEQLLEAPARQEE 957
 Db 1082 qikeleelrqtqkqlkeqlake 1104
 RESULT 14
 AAY49936
 ID AAY49936 standard; protein; 2101 AA.
 AC AAY49936;
 XX
 DT 01-FEB-2000 (first entry)
 DE Human NUMA protein #1.
 XX
 KW Human; NUMA; intranuclear protein; mitosis; segregation; DNA-PK-cs;
 KW DNA-activated protein kinase; catalytic subunit; PAPP; detection;
 KW nuclear NAD+-ADP-ribosyltransferase; autoantigen; immune response;
 KW autoimmune disease; cancer; type I diabetes mellitus; thyroiditis;
 KW myasthenia gravis; primary biliary cirrhosis; rheumatoid arthritis;
 KW systemic lupus erythematosus; polymyositis; dermatomyositis;
 KW Sjogren's syndrome; scleroderma; graft-vs-host disease.
 XX
 OS Homo sapiens.
 XX
 PN W0953757-A1.
 XX
 PD 28-OCT-1999.
 XX
 PF 22-APR-1999; 99WO-US08774.
 XX
 PR 22-APR-1998; 98US-0082643.
 XX
 PA (MERI) MERCK & CO INC.

PA (UYJO) UNIV JOHNS HOPKINS.
 XX (MERI) MERCK FROSST CANADA INC.
 PI Thornberry N, Rosen A, Casciola-Rosen L, Andrade FA, Nicholson D;
 XX Roy S;
 DR WPI; 2000-013162/01.
 XX Autoantigenic fragments useful for diagnosis, treatment of autoimmune
 PT diseases and cancer -
 PS Claim 4; Fig 9; 88pp; English.
 XX
 CC The present invention describes autoantigenic fragments (I) produced by
 CC the action of a lymphocyte granule enzyme (II) on isolated cells
 CC containing an autoantigen. (I) is useful for prophylactic and therapeutic
 CC treatment of an autoimmune disease (AI) such as type I diabetes mellitus,
 CC thyroiditis, myasthenia gravis, primary biliary cirrhosis, systemic lupus
 CC erythematosus, rheumatoid arthritis, polymyositis, dermatomyositis,
 CC Sjogren's syndrome, scleroderma and graft-vs-host disease. (I) produced
 CC by the action of (II) on cells isolated from target tissue is
 CC administered to a patient for tolerizing to the presence of (I). As a
 CC therapeutic treatment, isolated (I) associated with AI condition is
 CC contacted with serum of the patient containing autoantibodies (AAB)
 CC against (I), under condition to allow the in vivo binding of AAB to (I)
 CC and a portion of the AAB is removed from the serum of the patient. (I)
 CC derived from malignant cell is administered to a patient for stimulating
 CC an immune response against malignant cells to treat cancer. (I) is also
 CC useful in assays for assessing the presence or absence of an AI condition
 CC in a patient. Such assays involve detecting AAB in a sample by using (I)
 CC or detecting (I) in a sample by using antibodies that specifically binds
 CC to a cryptic epitope of (I). The present sequence represents the human
 CC NUMA protein, which is a specifically claimed autoantigen from the
 XX present invention.
 SQ Sequence 2101 AA;

Query Match 5.0%; Score 259; DB 21; Length 2101;
 Best Local Similarity 19.9%; Pred. No. 1.9e-11;
 Matches 208; Conservative 162; Mismatches 377; Indels 296; Gaps 39;
 QY 4 LCRRDSALTALDEETLWEMMESHRIRVRCICPSRLTPYLROAKVLCOLDEEVLHSPRL 63
 Db 263 llnekqaasplekeleelrdkn-----eslmrlhetlkkqcdlkteksqmdrk 312
 QY 64 TNSAMRAGHLDDLKTRGKNGAIAF-LESUKFHNPDVYTLVTLGTPDVPVDFSNFSGLMETS 122
 Db 313 inqlse-----engdisfklrefashlqlqldalneeltechskatgewlekqa 360
 QY 123 KLTECLAGAGISLOEELNQEKQKEVLLRRCQQLQELHGLAETRAEGLHOLEADHSHMKR 182
 Db 361 qlekeisaa---lqdkkcllee-kneillqgkisleenis-----qlqdnppqekg 406
 QY 183 EVSAHFHEVLRLKDEMLSL-----LHYSNALQEKELAAASRCRSLQELYL-- 228
 Db 407 evlgdvlqltqlqeaatlaanntqlgarvemletergqqaekllaerghfeekqqlss 466
 QY 229 -----LXQELQRANWVSCELELQEQSLRT-----ASDQESGDEELNRLK 268
 Db 467 litdlqssisnlsgakeeleqasqahgarltaqvasitseltlnatiqqq-dqelaglk 525
 QY 269 EE-----NEKLSLTFSLAEKDILOSLDEARSGRQELVERIH 306
 Db 526 qqakekgaqlaqltqqqeqasqglrhqveqlssslkqk---edqqlkev-aeqgeatrqdh 581
 QY 307 SLRERAAAEQRQEQWEEKQTLQFQKSMACQLYREKVNALQAQVCELOKQERDQAYS 366
 Db 582 a-qqlatsae-ereaslrerdaalkqle-----alekekaaklei-----lqqqlqvane 629
 QY 367 ARDSAQREISQSLVEKDSLRQVFE-----TDQVCELRTLRLQRLQAEPGP 412
 Db 630 ardsaqsvtqgarekelskrveelqacvetarqeqheaqgvaeeleqlrseq----- 684


```
QY      413 VLKQEARTREPCPREQRULVRMHAICPRDDSDCLSVSTESQLLSDSLASATSSRELVDSEFR   472  
       :|! ! | !! |  
Db     685 --qkatekervaqekdql-----geqlakelskvtkgseleeki    723  
  
QY      473 SSSPAPPSQQSLYKKVAEDEGPEPHSFSCLEIPEDGPCALPGAKAGDPHLDVELLDTAD    532  
       :|! ! | !! |  
Db     724 raadaleedqrctiselkae---tsrlvegknrkerkeleeeragrkglearlirlqe    776  
  
QY      533 LPOLESSLPVPSPGRILDVESGVLMRRRPARILLISOVTMLAFQDALLLEISIVIGGNLTG    592  
       || | | :|| | :|| | :|| | :|| | :|| | :|| |  
Db     777 ahqaete-----virrelaaemaqhct--aeseeceql-----    806  
  
QY      593 IFTHRTPGSAODMALRPCTOVMVDVEASEPLFKAVLEDITLEEAVGLLRVDGFCCFL    652  
       :|! ! | !! | :|| | :|| | :|| | :|| | :|| | :|| |  
Db     807 -----vkevaawrdgye----dsqqeaegyamfgeq-----lm    836  
  
QY      653 SVKVNTDGYYKRLLQDLKAUVTGSDFYIRVNLAAMEGRAKGELOVHCNEVLHVTDWTFQG    712  
       :|| |: || | :|| | :|| | :|| | :|| | :|| | :|| |  
Db     837 tlkeecekargelqeakekva-gieshesiqisrqnklael--han-----    880  
  
QY      713 CGCWHHRNVSYTMKDHTAHTGITPNYSRAOOOLIALIODWTQCOTTVTRKPSSGGGPQKL-    771  
       :|! ! | !! | :|| | :|| | :|| | :|| | :|| | :|| |  
Db     881 ----laraigqvgevrackladlstlqekmaatskevaretlvrk--sgeqgetas    934  
  
QY      772 RIYSMBKAKASPRLRSFDRGOLDPSRME--GSSTCFWAESCLTLVPYLWWHPHRRAPR    828  
       :|! ! | !! | :|| | :|| | :|| | :|| | :|| | :|| |  
Db     935 relvkepar-----gdrrpwieeqgrqfc-----stga    965  
  
QY      829 PVLILPRAVGKILSELKCLOGKFACKLAYLSQEEYEAWMSORGDIIOEGEVSGRCRWTR    888  
       :|! ! | !! | :|| | :|| | :|| | :|| | :|| | :|| |  
Db     966 alcamereaeagmgne-----lerlrallmesggqqegrgqerevarltqerg-----r    1015  
  
QY      889 HAVESLMEXNTHALLDVQDSVCTUHLRMDIFFI--VIHSVYNEMKAX-----KLK-----    936  
       :|| | :|| | :|| | :|| | :|| | :|| | :|| | :|| |  
Db     1016 agadlaaleakaaraelemrlqnalnegrvefatlqaalahaitekgdkgelakirglesa    1075  
  
QY      937 --KGLORLGTSBEQQLLEAAARQEE    957  
       || | | :|| | :|| | :|| | :|| | :|| | :|| |  
Db     1076 qikeeleeltqrtkvglkeqliakte    1098
```

| | | |
|----------|--------------------------|--|
| RESULT | 15 | |
| AB662423 | | |
| ID | AB662423 | standard; Protein; 1916 AA. |
| XX | XX | |
| AC | ABB62423; | |
| XX | XX | |
| DT | 26-MAR-2002 | (first entry) |
| XX | XX | |
| DE | Drosophila melanogaster | polypeptide SEQ ID NO 14061. |
| XX | XX | |
| KW | Drosophila; | developmental biology; cell signalling; insecticide; |
| KW | pharmaceutical. | |
| XX | XX | |
| OS | Drosophila melanogaster. | |
| XX | XX | |
| PN | WO200171042-A2. | |
| XX | XX | |
| PD | 27-SEP-2001. | |
| XX | XX | |
| PF | 23-MAR-2001; | 2001WO-US09231. |
| XX | XX | |
| PR | 23-MAR-2000; | 2000US-191637P. |
| PR | 11-JUL-2000; | 2000US-0614150. |
| XX | XX | |
| PA | (PEKE) | PE CORP NY. |
| XX | XX | |
| PI | Venter JC, | Adams M, Li PWD, Myers EW; |
| XX | XX | |
| DR | WPI; | 2001-656860/75. |
| DR | N-PSDB; | ABL06526. |
| XX | XX | |

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX Disclosure; SEQ ID NO 14061; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABBS7737-ABBS7072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1916 AA;

| | | | | |
|---------------------------|--------|--|-------------|--------------|
| Query Match | 5.08; | Score 258; | DB 22; | Length 1916; |
| Best Local Similarity | 22.1%; | Pred. No. 2e-11; | | |
| Matches 133; Conservative | 97; | Mismatches 209; | Indels 164; | Gaps 27; |
| Qy | 474 | SSPAPPQOQSLYKRVAEDFGEPEWFSFSCLEIPBPGDPCALPGAKAGDPHLDY----- | 525 | |
| Db | 1402 | nsrpprnslfp-----lpmqqaqstprgsraplshqskdqsfds | 1445 | |
| Qy | 526 | -----ELDDTADLPQLESSLPVSPGRLDVSESGVLMRRRPARRI-----LSQV | 569 | |
| Db | 1446 | lengdsissqdmppsaattttt-----asatstvydeekpalspppasvpaetlryv | 1500 | |
| Qy | 570 | TMLAFOGDALLEQISVTGGNLTGTFIHRVTPGSAADQMALRPGQIVMVDYEASEPLFKA | 629 | |
| Db | 1501 | tlhmdksknl--gklfggnkvglyvhdvavgspsdhagirkgqdl--leyngvd----- | 1551 | |
| Qy | 630 | VLEDTTLEAAGVLLRRYVDGFCCLSVKVNTDGYKRLLDQ-----LEAKVATSGDSFYIRVNL | 685 | |
| Db | 1552 | -lsgvtaeqaaneiskl-----tdvtmivqnlhtlkqikdepqdsfyrvgf | 1599 | |
| Qy | 686 | AMEGRA-KGELOVHCNEVLHVDTFMFOGC-GCWAHRAHNSYTMKDTAAH-----GTINYS | 739 | |
| Db | 1600 | drtgelneddlrfvdkdevlyvndvfnvgtfglrawkl-----danghrkecdilpsqm | 1653 | |
| Qy | 740 | RAOQQL-----IALIQMTOQCTYTRPSSGGPOKLRIVSMDX---AKASPLRLSF--- | 788 | |
| Db | 1654 | kveeelsrgevvdcdtgtarrgtsaarsffrkkngrssrdstelasfntqlsfpd | 1713 | |
| Qy | 789 | -----DEGQLDPSRMESSTCFWAESCLTILVPTLVWPHRPAPRPVLLVPRAVGKILS | 842 | |
| Db | 1714 | lgllndggaisrverllds-----prrrpfvllgp-----ll | 1747 | |
| Qy | 843 | E-----KLCLLQGF-----KKCLAEYLSQBEYEAWSORGDIIOBEGSVGGRCW | 885 | |
| Db | 1748 | eclmrltidfnslkficevtamdcseqameegikenifdvyrgrnkfe-----ct | 1799 | |
| Qy | 886 | VTRHAVESLMKENTHALLDVOLDVSCTILHRMDIPTPIVHVSNEMKAKKLKGQIRGTS | 945 | |
| Db | 1800 | tveaisnackndrrhcdlvdaisaverlqrlqlyplvl--llrfksakql-rdirdfgtd | 1856 | |
| Qy | 946 | E-----EOLLEAAROEGLDRAPCLYSSLAPDGSWDLCDLLSCVRQAIADKQKVVWT | 999 | |
| Db | 1857 | kisakaakemyeramkletdykq---yisavigpvs-ikhmctcqikdavedkedqkllw | 1911 | |
| Qy | 1000 | EQS 1002 | | |
| Db | 1912 | pvs 1914 | | |

Search completed: June 13, 2002, 09:23:52
Job time: 235 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2002, 09:21:47 ; Search time 17.94 Seconds
(without alignments)
1366.962 Million cell updates/sec

Title: US-09-767-215-2

Perfect score: 5149
Sequence: 1 MGELCRDSALTALDETLW.....VROAIADEQKVVWVTEQSPR 1004

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------------|--------------------|
| 1 | 262 | 5.1 | 2101 | 1 US-08-466-390-4 | Sequence 4, Appli |
| 2 | 262 | 5.1 | 2101 | 1 US-08-470-950-4 | Sequence 4, Appli |
| 3 | 262 | 5.1 | 2101 | 1 US-08-467-781-4 | Sequence 4, Appli |
| 4 | 262 | 5.1 | 2101 | 1 US-08-195-487-4 | Sequence 4, Appli |
| 5 | 262 | 5.1 | 2101 | 2 US-08-483-924-4 | Sequence 4, Appli |
| 6 | 262 | 5.1 | 2101 | 4 US-09-452-294-1 | Sequence 1, Appli |
| 7 | 262 | 5.1 | 2101 | 5 PCT-US93-06160-4 | Sequence 4, Appli |
| 8 | 238 | 4.6 | 3248 | 1 US-08-353-700-1 | Sequence 1, Appli |
| 9 | 238 | 4.6 | 3248 | 5 PCT-US95-16216-1 | Sequence 1, Appli |
| 10 | 237 | 4.6 | 2482 | 1 US-08-328-254-6 | Sequence 6, Appli |
| 11 | 216.5 | 4.2 | 1388 | 2 US-08-685-576-1 | Sequence 1, Appli |
| 12 | 210.5 | 4.1 | 976 | 4 US-09-104-324B-4 | Sequence 4, Appli |
| 13 | 205 | 4.0 | 816 | 2 US-08-533-306A-6 | Sequence 6, Appli |
| 14 | 205 | 4.0 | 816 | 2 US-08-742-923A-6 | Sequence 6, Appli |
| 15 | 205 | 4.0 | 885 | 2 US-08-533-306A-4 | Sequence 4, Appli |
| 16 | 205 | 4.0 | 885 | 2 US-08-742-923A-4 | Sequence 4, Appli |
| 17 | 204.5 | 4.0 | 1354 | 3 US-08-685-871-2 | Sequence 2, Appli |
| 18 | 200.5 | 3.9 | 1388 | 2 US-08-685-576-4 | Sequence 4, Appli |
| 19 | 197 | 3.8 | 576 | 2 US-08-533-306A-2 | Sequence 2, Appli |
| 20 | 197 | 3.8 | 576 | 2 US-08-742-923A-2 | Sequence 2, Appli |
| 21 | 197 | 3.8 | 1886 | 4 US-08-938-105-3 | Sequence 3, Appli |
| 22 | 195 | 3.8 | 1090 | 4 US-09-085-199B-5 | Sequence 5, Appli |
| 23 | 194 | 3.8 | 1388 | 4 US-09-572-191-2 | Sequence 2, Appli |
| 24 | 192.5 | 3.7 | 756 | 4 US-09-085-199B-9 | Sequence 9, Appli |
| 25 | 190 | 3.7 | 914 | 4 US-09-085-199B-4 | Sequence 4, Appli |
| 26 | 189 | 3.7 | 1898 | 1 US-08-056-200-94 | Sequence 94, Appli |
| 27 | 189 | 3.7 | 1898 | 2 US-08-800-644-94 | Sequence 94, Appli |

28 189 3.7 1939 4 US-09-310-187A-1 Sequence 1, Appli
29 186.5 3.6 955 1 US-08-006-676B-1 Sequence 2, Appli
30 186.5 3.6 955 1 US-08-282-845-2 Sequence 1, Appli
31 186.5 3.6 955 5 PCT-US94-00324-1 Sequence 1, Appli
32 186 3.6 683 6 5210183-3 Patent No. 5210183
33 182.5 3.5 955 2 US-08-428-414A-3 Sequence 3, Appli
34 181 3.5 1618 1 US-07-853-913-4 Sequence 4, Appli
35 176.5 3.4 900 2 US-08-630-822A-62 Sequence 62, Appli
36 176.5 3.4 900 2 US-09-005-069-62 Sequence 62, Appli
37 176 3.4 1162 2 US-08-728-323A-2 Sequence 2, Appli
38 175 3.4 835 2 US-08-968-751-4 Sequence 4, Appli
39 175 3.4 1068 4 US-09-085-199B-11 Sequence 11, Appli
40 174.5 3.4 712 2 US-08-468-576B-17 Sequence 17, Appli
41 174.5 3.4 712 2 US-08-468-579B-17 Sequence 17, Appli
42 174.5 3.4 712 3 US-08-468-577B-17 Sequence 17, Appli
43 173.5 3.4 1375 4 US-09-722-139-2 Sequence 2, Appli
44 166.5 3.2 1312 2 US-08-687-080-51 Sequence 51, Appli
45 165.5 3.2 829 1 US-07-670-611-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-466-390-4
; Sequence 4, Application US/08466390
; Patent No. 5686562
; GENERAL INFORMATION:
; APPLICANT: TOUTKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,390
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-466-390-4

Query Match 5.1%; Score 262; DB 1; Length 2101;
Best Local Similarity 19.9%; Pred. No. 5.7e-13;
Matches 208; Conservative 163; Mismatches 376; Indels 296; Gaps 39;
QY 4 LCRDSALTALDETLWEMSHRRIVRCIPSRLLPYLRQAKVLCOLDEEVLHSPRL 63
DB 263 LLNEKQAASPLEPKELEURDKN-----ESLTMRLHETLKCQODLUKERSQMDRK 312


```
Db 935 RELVKEPARA-----GDROPEWLEEQGRQFC-----STQA 965
QY 829 PVLVPRVAGKILSEKLCLLQGFKKCLAEYLSOEYEAWSQRGDIIOEGVSGRCWVTR 888
Db 966 ALQAMEREAEQMGNE-----LERLRAALMESQGOQOERGOQEREVARLTQERG-----R 1015
QY 889 HAVESLMENKTHALLDVQSDVCTLHRMDIFPI--VIHVSVNEKMAK-----KLK----- 936
Db 1016 AQADLALAEKAAAELEMLQNALNEQREVFATLQEAHALAHALTEKEGKDOELAKIRGLEAA 1075
QY 937 --KGLQRLGTSEBOLLEAAQREE 957
Db 1076 QIKELEELRQTVKQLEQAKKE 1098

RESULT 4
US-08-195-487-4
; Sequence 4, Application US/08195487
; Patent No. 5783403
; GENERAL INFORMATION:
; APPLICANT: TOKUATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,487
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/901,701
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-195-487-4

Query Match 5.1%; Score 262; DB 1; Length 2101;
Best Local Similarity 19.9%; Pred. No. 5.7e-13;
Matches 208; Conservative 163; Mismatches 376; Indels 296; Gaps 39;

QY 4 LCRRDSALTALDEETLWEMWESHRHVRICPSRITPYLRQAKVLCQDDEEVLSHRL 63
Db 263 LNEKQAASPLEPKELELRDN-----ESUTMRLEHTLKQCQDLKTEKQMDRK 312
QY 64 TNSAMRAGHLDDLKTRGNATAP-LESLEKFNPDVYTLVTGLQPDVDFSNFSGLMETS 122
Db 313 INOLSE-----ENGDSUFLKREFASHLQQLDALNELTEHSKATQEWLEKOA 360
QY 123 KLTECLAGATGSLQELNQEKGKQVLLRRCCQQLQELHGLAETRAEGLHQLHEADHSMKR 182
```

```
Db 361 QLEKELSA--LQDKKCLEE-KNEILQGLKSLQEEHLS-----QLQNPPOEKG 406
QY 183 EVSAHFHEVLRLQKDEMLSL-----LHYSNALQEKELAAACRCSLSQEEVL-- 228
Db 407 EVLGDVLQETLQKQAEALANNTQLOARVEMLETERGQOQEAKLAEERHFEKEQQLSS 466
QY 229 -----LKOELQRANMVSSCELELOFQSILRT-----ASDQSGDEELNRLK 268
Db 467 LITDLOSSISNLQAKKELEQASQAHGARGLTQAASILTSELTILNATIOQO-DOELAGLK 525
QY 269 EE-----NEKRLSLTSLAEOKIOLESLDARSROELVERIH 306
Db 526 QOAKEKQAOAQTLOQOQEQASQGLRHQVEOLSSSLKQK-----EQOLKEV-AEKQEAQRDH 581
QY 307 SLRERAVAAERQREYWEKEQTLLOFQKSKMACOLYREKVNALQAOVCLEKQERDOAYS 366
Db 582 A-QOLATAAB-EREASLURERDAALKQLE-----ALEKKAALKEI-----LQOOLQVANE 629
QY 367 ARDSAQREISQSLVEKDSLRQVFEL-----TDQVCELRTOLRQLQAEPPG 412
Db 630 ARDSAQTSVTQAOREKAELSRKVEELQACVETARQEOHEAQAOVAELEQLRSEQ----- 684
QY 413 VLQOEARTEPCPREKQRLVRMHAICPRDDSDCSLVSTESQLLSLSATSSRELVDSEFR 472
Db 685 ---QKATEKERVAKQDKL-----QEQLQALKESLKVTKGSLEEEKR 723
QY 473 SSSPAPPQSQSLYKRVAFDGEPEWSEFSSCLEITPEGDPGALPGAKAGDPHLDYELDTAD 532
Db 724 RAADALEEQORCISELKAEE-----TRSLVEHQHKKERKEERAGKGLERLQLQGE 776
QY 533 LPQLESLOPVSPGRLDVSESGVLMRRRRPARRILSOVTMLAFQGDALLEGISVIGNLTG 592
Db 777 AHQAETE-----VLRRELAEAAMAQHT-----AESCEQL----- 806
QY 593 IFTHRTVPGSAADOMALRPGCTQIWMVDYEAASEPLFKAVLEDTTLEEAAGVLLRRVDGFCCL 652
Db 807 -----VKEVAAMRDGYE-----DSQOEEAQYGAAMFOEQ-----LM 836
QY 653 SVKVTNDGYKRLQLDLQLEAKVATSGDSFYIRVNLAMEGRAGKEQLQVHCNEVLHVTDTMFQ 712
Db 837 TLKEECEKARQELQEAKEVA--GIESHSELOISRQONKLAEL--HAN----- 880
QY 713 CGCWHAHVNSYTMKDTAAHGTTIPNYSRAOQIILQDMTQOCTVTRKPSGGPKLV- 771
Db 881 ----LARAQQOQOEKVRQAKLADLLSTLQEKMAATSKEVARLETIVRK--AGEQOQETAS 934
QY 772 RIVSMDKAKASPLRLSFDRCQLDPSRME---GSSTCFWAESCLTLPVYTLVWPHRPARPR 828
Db 935 RELVKEPARA-----GDROPEWLEEQGRQFC-----STQA 965
QY 829 PVLVPRVAGKILSEKLCLLQGFKKCLAEYLSOEYEAWSQRGDIIOEGVSGRCWVTR 888
Db 966 ALQAMEREAEQMGNE-----LERLRAALMESQGOQOERGOQEREVARLTQERG-----R 1015
QY 889 HAVESLMENKTHALLDVQSDVCTLHRMDIFPI--VIHVSVNEKMAK-----KLK----- 936
Db 1016 AQADLALAEKAAAELEMLQNALNEQREVFATLQEAHALAHALTEKEGKDOELAKIRGLEAA 1075
QY 937 --KGLQRLGTSEBOLLEAAQREE 957
Db 1076 QIKELEELRQTVKQLEQAKKE 1098

RESULT 5
US-08-483-924-4
; Sequence 4, Application US/08483924
; Patent No. 5882876
; GENERAL INFORMATION:
; APPLICANT: TOKUATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
```

NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,924
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ. EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-483-924-4

Query Match 5.1%; Score 262; DB 2; Length 2101;
Best Local Similarity 19.9%; Pred. No. 5.7e-13;
Matches 208; Conservative 163; Mismatches 376; Indels 296; Gaps 39;

QY 4 LCRDSDALTALDEETLWEMMESHRIRVRCICPSRLTPYLQAKVLCQDDEEVLHSPRL 63
DB 263 LLNEKQAASPLEPKELEELRDKN-----ESLTMRLHETLKQCQDLKTEKSDMRK 312
QY 64 TNSAMRAGHLDDLKTRGKNGAIAF--LESKLFHNPDPVYTLVGLQPDVDFSNFSGLMETS 122
DB 313 INQLSE-----ENGDSLKREFASHLQQLDALNELTEHESKATQEWLEKQA 360
QY 123 KTECLAGAGISLOEELNEKGKQVLLRRCCQLOQLHGLAETRAEGLHOLEADHSMRKR 182
DB 361 QLEKELSA--LQDKKLEE--KNEILQGLKSLQLEHLS-----LQDNPPQEK 406
QY 183 EYSAHFHEVLRKDEMLSL-----LHYSNALQEKELAAASRCRSLEELYL-- 228
DB 407 EVLGDVQLQETLKQEAATLAANNLTQARVEMLETERGQEQAKLLAERGHFEKQQLSS 466
QY 229 -----LQELQANMYSSCELELQEQSLRT-----ASDQESGDEELNRUK 268
DB 467 LITDLOSSISLSQAKEELEQASQAHGRLTAQVSLTSELTLNATIOQQ--DQELAGLK 525
QY 269 EE-----NEKLSLTFSLAEKDILEQSLDEARGSRQELVERTH 306
DB 526 QAKAQKQAQLQTLQEQEQASGLRHQVQLSSLKQK---EQQLKEV-AEKQEAATRDQH 581
QY 307 SURERAVAAERQEQWEEKEQTLLOFKQSKMACQLYREKVNALQAVGCEQLKERDQAYS 366
DB 582 A-QQLATAAE-EREASLREDAALKLE-----ALEKEKAALKLEI---LQQLQVANE 629
QY 367 ARDSAQRETSQSLVEKDSLRQVFEI-----TDQVCELRQLQOLQAEPPG 412
DB 630 ARDSAQTSVTOQREKAELSRSVEELQACVETARQEQHEAQVAELEQLRSEQ----- 684
QY 413 VLKQEAETREPCREKQRLVRMHAICPRDSDCSLSVSTESQLLSLSSRELVDSEFR 472
DB 685 ---QKATERVAQEKDQL-----QBQLQALKESLKVTKGSLEEKR 723

QY 473 SSSPAPPSSQSLYKRVAEFDGPEPMWSFSCLTEIPGDPGALPKAGADPHLYELDTAD 532
DB 724 RAADALEEQORCISELKAB-----TRSLVEQHKRKEKEERAGRAGRLQLQGE 776
QY 533 LPQLESSLPVSPGRDLVSESGLVMRRRPARRILTSQVTLMAFGDALLQISVIGNLTC 592
DB 777 AHOAETE-----VLRRELAEMAAOHT-----AESECBOL----- 806
QY 593 IFIHRVTPGSAADOMALRPGTQIVMVDYEAASEPLFKAVLEDTTLEEAVALRRVDFCCL 652
DB 807 -----VKEVAARRDGYE-----DSQEEAQYCAMFQEQ-----LM 836
QY 653 SVKVNTPGYKRLQLDLQLEAKVATSGDSFYIRVNLAMGRAGKELQVHCNEVHLVHTDMFQ 712
DB 837 TLKECEKARQELQEAKEKVA--GIESHSELQISROONKLAEL--HAN----- 880
QY 713 CGCWAHRRVNSYTMKDTAAHGTPINYSRAQQQLIALIQDTQOCTVTRKPSGGPQKLV- 771
DB 881 ----LARALQOQVOEKEVRAQKLAADDLSTLOEKMAATSKVARLETIVRK--AGEQOETAS 934
QY 772 RIVSMDKAKASPLRLSFDRLQDPPSRME--GSSTCFWAESCLTLVPTYTLVWPHRPARPR 828
DB 935 RELVKEPARA-----GDRQPEWLEEQGQRFQ-----STQA 965
QY 829 PVLLPRVAVGKILSEKLLQGLFKKCLAEYLSQEEYEAMWSQRGDIIOEGEVSGRCWVTR 888
DB 966 ALQAMEREAEQMGNE-----LERLRAALMESQGOQOQEQEREVARLTQERG-----R 1015
QY 889 HAVESLMKENTHALLDVQLDSVCTLHRMDIFPI--VIHVSVNEKMAK-----KLK----- 936
DB 1016 AQADLAKAARAELMRLONALNEQRFATLOEALAHALTEKQKQOQELAKRLGLEAA 1075
QY 937 --KGLQRLGTSEEQLEAAARQEE 957
DB 1076 QIKELEELRQTVKQLKEQAKKE 1098

RESULT 6
US-09-452-294-1
; Sequence 1, Application US/09452294
; Patent No. 6287790
; GENERAL INFORMATION:
; APPLICANT: Lelievre, Sophie
; APPLICANT: Bissell, Mina
; TITLE OF INVENTION: UTILIZATION OF NUCLEAR STRUCTURAL PROTEINS FOR TARGETED
; TITLE OF INVENTION: THERAPY AND DETECTION OF PROLIFERATIVE AND
; TITLE OF INVENTION: DIFFERENTIATION DISORDERS
; FILE REFERENCE: IB-1454- Sequence Submittal
; Patent No. 6287790
; CURRENT APPLICATION NUMBER: US/09/452,294
; CURRENT FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: 60/110,420
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2101
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-452-294-1

Query Match 5.1%; Score 262; DB 4; Length 2101;
Best Local Similarity 19.9%; Pred. No. 5.7e-13;
Matches 208; Conservative 163; Mismatches 376; Indels 296; Gaps 39;

QY 4 LCRDSDALTALDEETLWEMMESHRIRVRCICPSRLTPYLQAKVLCQDDEEVLHSPRL 63
DB 263 LLNEKQAASPLEPKELEELRDKN-----ESLTMRLHETLKQCQDLKTEKSDMRK 312
QY 64 TNSAMRAGHLDDLKTRGKNGAIAF--LESKLFHNPDPVYTLVGLQPDVDFSNFSGLMETS 122
DB 313 INQLSE-----ENGDSLKREFASHLQQLDALNELTEHESKATQEWLEKQA 360


```

Db 2784 ---KEGVREIEAAYQLRLHEAEKKHQAALLDNTNKOYEVEIQYREKLTSEKELSSQKL 2840
QY 771 VRIVMDKAKASPLRLSFD-----RGQLDPSRMGSSSTCFWAESCLTLPVTVLWPHR 823
Db 2841 ----EIDLKSSKEELNLSLKATTOILEELKTKMD-----NLKYVNO 2879
QY 824 PARPRVLLVPRAVGKI-LSEKLC-LLOGFKKCLAEVLSQEEYEANWSQGDIIQGEVSG 881
Db 2880 LKKN-----ERAQGMKMLIKSKQLEEEKEILOKELSOLAQAQKQKTGTWMDTKVDE 2934
QY 882 GRCWVTRHAVESLMKENTHALLDVLDSVCTLHRMDIFFIVHVSNEKMAKKLKGQLO 941
Db 2935 LTTEI-KELKETLEETKEA--DEYLDKYCSL-----LISH-----EKLEKAKEM 2976
QY 942 LGTSEQLLEAARQEGDLDRAPCL 966
Db 2977 LETQVAHL--CSQOSKODSRGSPLL 2999

RESULT 9
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
PCT-US95-16216-1

Query Match 4.6%; Score 238; DB 5; Length 3248;
Best Local Similarity 19.4%; Pred. No. 1.3e-10;
Matches 214; Conservative 192; Mismatches 393; Indels 306; Gaps 48;

QY 2 GELCRDSALTALD-----EETLWEMESHRRIVRCICPSRLTPYLRAQVLCQDE 54
Db 2061 GELDTMSKTTALDQLSEKMKET--QELSHQSECLHCI-----QVAE 2102

```

```

QY 55 EEVLHSPRLTNSAMRAGHLDDLKTRGKNGAIAFLBSLKFHPNDVYTVLTGLOPDVDFSN 114
Db 2103 AEVKEKTELLQTL--SSDVELLKD-----KTH--LOEKLOSLEKD----- 2139
QY 115 FSLMETSKLTECLAGISLOEELNOEKOKGOKVLLRRCOOLOEHLG----- 161
Db 2140 ----SOALSUTKC-----ELENOIAQLNKEKELLKESLOARLESSEYKLNVSAL 2189
QY 162 -----LAETRAEGLHOLEADHSRMKREYSA-----HFHEVLR-----L 194
Db 2190 EAALVEKGEFALRLSTQEE-VHQLRGIEKLVRIEADKKQLHIAEKLKERENDSL 2248
QY 195 KDEMLSLSLHYNALQEKELAEASCRSLQBELLYLLQELQORANWSSCEL-ELQEOSLRT 253
Db 2249 KDKVENLERELQWSENOELVIIDAENSKAETVLTQIE--EMARSLKIFELDLVTLRS 2306
QY 254 ASDQ--ESGDEELNRLKEENKLSLTFSLAEKDILE-QSLDEARGSQBELVERIHSRE 310
Db 2307 EKENLTQIQEKGQGLSGLDKLLSSFKSLLEKEQAEIQIKEESKTAVEMLOQLKELNE 2366
QY 311 RAVAAERQREYWEKEQTL-----LQFQSKMACOLYREK----- 346
Db 2367 -AVALCGDQEIIMKATEQSLDPPIEBEHOLRNSIEKLRLARLEADEKKQLCVLOQLKESEH 2425
QY 347 -VNALQAQVCELOKQDQAYSARDSAQREISQSLVEKDSLRQVVFELTDOVCBLRTQLRQ 405
Db 2426 HADLLKGRVENLERELIARTNQEHAALEAENSKEGVEVLKAKIEGTMQ-----SLRG 2478
QY 406 LOAEPFVLKOEARTPCPREKORLVRMHAICPRDDSDCSLSVSSTESQLLSLSATSSR 465
Db 2479 LEDVVTIRSEKENLTNLOKEQERI-----SELEINSSFENILQE-----KEQ 2523
QY 466 ELVDSPRSSPAPPSSQSLYKRVADFGPEPWSFSSCLEIPEGDPCALPGAKADPHLDY 525
Db 2524 EKVMKESSTAMEMLQTLKELNERNVAALHNDQEA-----KAKEQNLSS 2569
QY 526 EL-----LDTRADLPQ-----LESS-----LQVSPGRLDVSESGVLMRRRPARIL 566
Db 2570 QVECLEKAQLOGLDEAKNNYIVLOSSYKGLQIEVEDGQKLEK-----KDEISRLK 2624
QY 567 SQVTMLAFQGDALLEQISVIGG-----NLTGFIHRVTPGSAADQMAIRPGT 613
Db 2625 NQIQ----DQELVSKLSQVEGEHQLWKEQNLRLNLT-VELEQKIQVLSKNASLODTL 2679
QY 614 QIVWVDYEAASEPLFKAVLEDTTLEEAVALRRVDGFCCLSVKVTNGYKRLLODLAKVA 673
Db 2680 EVLQSSYKNLE---NELELTMDK-MSFYEKVNKMTAKETELQREHMAQTAELQEE 2734
QY 674 TSGDSFYIRVNLAMEGRAGELQVHCNEVLHVTDTMFO-----GCGCWHARVNS 723
Db 2735 LSQE-----KNRLAGELQLLEIEIKSKDQKLELTENSELKSLDCHMKDQVE-- 2783
QY 724 YTMKDTAAHGTIPNYS-----RAQQOIALIQTDMTOQCTV-----TRKPSGGPKL 770
Db 2784 ---KEGVREIEAAYQLRLHEAEKKHQAALLDNTNKOYEVEIQYREKLTSEKELSSQKL 2840
QY 771 VRIVMDKAKASPLRLSFD-----RGQLDPSRMGSSSTCFWAESCLTLPVTVLWPHR 823
Db 2841 ----EIDLKSSKEELNLSLKATTOILEELKTKMD-----NLKYVNO 2879
QY 824 PARPRVLLVPRAVGKI-LSEKLC-LLOGFKKCLAEVLSQEEYEANWSQGDIIQGEVSG 881
Db 2880 LKKN-----ERAQGMKMLIKSKQLEEEKEILOKELSOLAQAQKQKTGTWMDTKVDE 2934
QY 882 GRCWVTRHAVESLMKENTHALLDVLDSVCTLHRMDIFFIVHVSNEKMAKKLKGQLO 941
Db 2935 LTTEI-KELKETLEETKEA--DEYLDKYCSL-----LISH-----EKLEKAKEM 2976
QY 942 LGTSEQLLEAARQEGDLDRAPCL 966
Db 2977 LETQVAHL--CSQOSKODSRGSPLL 2999

```


; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/685,576
 ; FILING DATE: 24-JUL-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 7-325129
 ; FILING DATE: 20-NOV-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 8-17150
 ; FILING DATE: 05-JAN-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 8-131206
 ; FILING DATE: 26-APR-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Bent, Stephen A.
 ; REGISTRATION NUMBER: 29,768
 ; REFERENCE/DOCKET NUMBER: 16887/843
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)672-5300
 ; TELEFAX: (202)672-5399
 ; TELEX: 904136
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1388 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-685-576-1

Query Match 4.2%; Score 216.5; DB 2; Length 1388;
 Best Local Similarity 19.6%; Pred. No. 2.3e-09;
 Matches 151; Conservative 129; Mismatches 292; Indels 199; Gaps 28;
 QY 6 RDSALFALDEE-TLWEMESHRRIVR--CICPSRLTPYLQKVLQCLDEEVLHSPR 62
 DB 476 RLEKVALEEEITLRNVESTLRQLERAKALQKNAEYORKA-----DHEADKRN 528
 QY 63 LTNSAMRAGHLDDLKTRGNGATF----- 88
 DB 529 LENDVNSLKQDLDLKKRNQNSQISTEKNVQLQQLQDQETNALLRTESDTAARLKRTOAES 588
 QY 89 ---LESKFHNPVY---TLVLTG-LQPDVDFSNFSGLMET-----SKLTCLAGAI 132
 DB 589 SKQIQOLESNNRDLQKNCLELAKLKEFIINLQSVLESERRDRTHGSEIINDLQGR 648
 QY 133 GSLOEELNQEKGQEVLLRRQ---QLOEHLGLAETRAEGL-----HOLEADHSRMKRE 183
 DB 649 SGLEEDVKNKG---ILLAKVELEKRLQERFTDLEKEKNMEIDMTYQLKVIQOOSLEQE 704
 QY 184 VSAHFHVLRLKDE---MLSLSHYSNALQKELAAASRCRSLOPELYLLKQELQRANNVS 240
 DB 705 ETEHKATKARLADKNNKIVETSEAKSEAMKEKLLSEERTLKQKVENLLLEAEKRSIL 764
 QY 241 SCELELOPQSLR-----TASDQESGDBELN----- 265
 DB 765 DCDLKQSQOKINELLKQKQDLNEDVRNLTKIEQETQKRCLTQNDLKMOTQOVNTLKMSE 824
 QY 266 -RLKEENEKLSLFTSLAEKDI-LEQSLDEARGSRQELVERIHSRLREAVAAERQRYW 323
 DB 825 KQLQENHLLKMKMSLEKQNAELRKERQDADGQMKELQDQL-----EASOYF 872
 QY 324 EKEQTLQFOKSKMACOLYREKVNALQAVCELOKQRQAYSAARDSAQREISQSLSEKD 383
 DB 873 STLYKT--QVRELKECEKEKTKLCKELQKQKQELQDERDSL-----AAQLEITLTKADSE 925
 QY 384 SLRRQVFEITQVCELRQLQOLQNEPGLVKQEARTEPCPREKQRLVRMHATCPRDSD 443
 DB 926 QLARSIAE-----BQYSDLEKEK---IMKELEIKENMARHKEQELT-----EKDA 966

QY 444 DCSLYSVSTESQLLSDLS--ATSSRELVDVSRSSPAPPSQQSLYKRVADDFCEEPWSPSS 501
 DB 967 TIASLEETNRTLTSDVANLANEKEELNNKLKE-----AQEQLSKLDEEI----- 1011
 QY 502 CLEIPGDPGALPGAKAGDPHLDYELLD-----TADLPOLESSLOQVSPG-----R 547
 DB 1012 -----SAAAIKAQFEKQQLLTERTLKTQAVNKLAIEIMNRKEPVKRGNDTDVPR 1058
 QY 548 LDVSESGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTGFIHRTVTPGSAADQM 607
 DB 1059 KEKENRKLHMLKSEREKLQ-QMIKYQKELNEMQAIAEESQIRIELQMTLDSKDSQDIE 1117
 QY 608 ALRPGTQIVWVDYEASEPLFKAVLEDTTLEBAVGLLRVVDGFCCLSVKYNVT 658
 DB 1118 QLRSQLQALHIGLIDSS--IGSGPGDTEADGGFP-ESRLEGLWSLPVRNNT 1165
 RESULT 12
 US-09-104-324B-4
 ; Sequence 4, Application US/09104324B
 ; Patent No. 6232460
 ; GENERAL INFORMATION:
 ; APPLICANT: T reci, Ozlem; Sahin, Ugur; pfreundschuh, Michael
 ; TITLE OF INVENTION: Methods For Diagnosis And Treating Cancers,
 ; TITLE OF INVENTION: And Methods For Identifying Pathogenic Markers In A Sample
 ; TITLE OF INVENTION: NO. 6232460mal Cells
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fulbright & Jaworski LLP
 ; STREET: 666 Fifth Avenue
 ; CITY: New York City
 ; STATE: New York
 ; ZIP: 10103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
 ; COMPUTER: IBM
 ; OPERATING SYSTEM: PC-DOS
 ; SOFTWARE: Wordperfect
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/104,324B
 ; FILING DATE: 25-June-1998
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/892,702
 ; FILING DATE: 15-July-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hanson, No. 6232460man D.
 ; REGISTRATION NUMBER: 30,946
 ; REFERENCE/DOCKET NUMBER: LUD 5491
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 318-3000
 ; TELEFAX: (212) 752-5958
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 976 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; US-09-104-324B-4

Query Match 4.1%; Score 210.5; DB 4; Length 976;
 Best Local Similarity 22.8%; Pred. No. 4.2e-09;
 Matches 119; Conservative 76; Mismatches 189; Indels 139; Gaps 20;
 QY 12 TALDEETLWEMESHRRIVRCICPSRLTPYLQAKV-----L 49
 DB 303 TKLQSENKOSIEKQHH-----LTRELEDIKYSQRSVSTQKALBEDLQIATKTI 352
 QY 50 COLDEEE-----VLHSPRLTNSAMRAGHLDLLKTRGNGAIAFLESKLFHNP 98
 DB 353 COLTEKETOMESENKARAHHSFVVFTEFTVCSLEELLRT-----QORLEKNEQ 404
 QY 99 VYTLVTGLQPDVDFSNFSGMLMETSCLTECLAGATGSLQELNQEKGQEVLLRRCQOL-- 156

| | | | | | |
|---------------------------|-----|--|-----------------------------------|----------------------|--------------|
| Best Local Similarity | | 20.8%; | Pred. No. 9.3e-09; | | |
| Matches 105; Conservative | | 105; | Mismatches 184; | Indels 110; Gaps | |
| QY | 40 | TPYLROAKVLCQDDEEVLHSP---- | RLTNSAMRAGHLDLLKTR---- | GKNGATAPLE--- | 90 |
| Db | 361 | TKALSTARAL----- | EEALEAKEELERTNKMKA- | EMEDLVSVSKDDVGKN-- | VHELEKSK 412 |
| QY | 91 | -SLKFHPNDVYTLVTGLQPDVDFSNFSGLMETSKLTCLAGATGSGLEELNQEKQEVYL 149 | | | |
| Db | 413 | RALETQMEENKTLLEELEDELQASDAKL----- | RLEVNMQAKQFPERDLQARDEQNE-- | | 466 |
| QY | 150 | LRCQOQLQELHGLAETPAEGLHQLEADHSRMKREVSAAHPHEVLVRKDEMLSLSHYSNAL 209 | | | |
| Db | 467 | -EKRRQLQRLQHEYT----- | ELEDERNERAAAAAKK-- | KLEGDKLKDLQADSAI 515 | |
| QY | 210 | QEKLEAASRCSQKSEELYLLKQELQ----- | RANWVSCELELEQEQL 251 | | |
| Db | 516 | KGREEATKQLRKLOAQMKQDFQRELEPARASDEIFATAKENKAKSLADMLQLEDLIA 575 | | | |
| QY | 252 | RTASDQSGDEELNRKEE----- | NEKLSLTFSLAEKDILQESLOEARGSRQE 300 | | |
| Db | 576 | AAERAKQAOALEKEELAEELASLSSGRNALQDEKRRLEARIAQ---- | LEEELEBEEQGNMEA 632 | | |
| QY | 301 | LVERIHSLSRRA----- | VAAERQREQYWEKEQTLLQFKSKMACQL----- | YR 344 | |
| Db | 633 | MSDRVKATQOABQLSNELATERSTAKNESAQQ-- | LERQNKEKLSKLHEMEGAVKSKPE 691 | | |
| QY | 345 | EKNVALQAQVCEQLKQRDQAYSARDSQAORISQSVLEKDSLRQVPELTDVCELTQRL 404 | | | |
| Db | 692 | STIAALEAKTAQLEEQVEQAREKQAA---- | TKSLQAKOKKLEILLQVEDERKMAEQVK 747 | | |
| QY | 405 | QLQAEPPGVILKQARTPEPCPREKQRLVRMHATCPRDDSDCVLSVSTESQLSDL--SATS 463 | | | |
| Db | 748 | EQAEKGNARVQLKRLQLEEAEEESQR----- | INANRRKLQRELDEATE 790 | | |

```

CORRESPONDENCE ADDRESS:
ADDRESS: Harness, Dickey & Pierce, P.L.C.
STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI
COUNTRY: USA
ZIP: 48303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

```

Query Match 4.0%; Score 205; DB 2; Length 816;

```

RESULT 14
US-08-742-923A-6
; Sequence 6, Application US/08742923A
; Patent No. 5869611
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/742.923A
; FILING DATE: No. 5869611member 1, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smitch, DeAnn F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-00869DVC
; TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-742-923A-6

Query Match 4.0%; Score 205; DB 2; Length 816;
Best Local Similarity 20.8%; Pred. No. 9.3e-09;
Matches 105; Conservative 105; Mismatches 184; Indels 110; Gaps 20;

QY 40 TPYLQAKVLCQDEEVLHSP---RLTNSAMRAGHLLDLLKTR---GKNGAIAFLF--- 90
Db 361 TKALSARL-----EALAEKEELERTNKLKA-EMEDLVSSKDDVGKN--VHELEKSK 412

QY 91 -SLKFHPDVTYLTGLOPDVDFNSGLMETSCLTECLAGATGSLQEEINQKQKEVL 149
Db 413 RALETOMEENKTOLEEELEQASEDAKL----RLEVNMQALKGQFERDLQARDEQNE-- 466

QY 150 LRCQQLQELHGLAETRAEGLHLEADHSRMKREVSAPHFHVLRLKDEMLSLSLHYSNAL 209
Db 467 -EKRRQLQRLQHEYT-----ELEDNERALAAAKK--KLEGDLKDLQADSAI 515

QY 210 QEKELAAASRCRSLQEEYLLKQELQ-----RANMVSCELELQBSL 251
Db 516 KGREEAIKQLRKLAQAKMDFORELADARASRDEIFATAKENKKAKSLEADLMQLEDLA 575

QY 252 RTASDQSGDEELNRLKEE-----NEKLSLTFSLAEKDILEQSLDEARGSRQE 300
Db 576 AAEPRARQAOLEKEELAEELASSLSGRNALQDEKRRLEARIAQ---LEELEEEQGNMEA 632

QY 301 LVERIHSRLRA-----VAAERQREYWEKEQTLLQFQSKMACOL-----YR 344
Db 533 MSDRVKATQQAQLSNELATERSTAGKNESARQQ--LERQNKELRSLKHEMGAVKSKFK 691

QY 345 EKVNALQAQVCELOKQERDQAYSARDSAQREISQSLVEKDSLRQVFLTDQVCELRQLR 404
Db 692 STTAALEAKIAQLEEQVEQAREKQAA---TKSLQKQKKKLEILLQVEDERKMAEQYK 747

QY 405 QLOAEPPGVLKQEARTEPCPREKQRLVRMHAICPRDDSCSLVSTESLSDLSL-SATS 463
Db 748 EQAEKGNARVQLKRLQLEEAEEESQR-----INANRRKLQRELDTE 790

QY 464 SREL----VDSFRSSSPAPPSSQOS 483
Db 791 SNEAMGREVNALKSLKLRGPPPPQET 814

RESULT 15
US-08-533-306A-4
; Sequence 4, Application US/08533306A
; Patent No. 5837457
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533,306A
; FILING DATE: September 25, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, DeAnn F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-00869COB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 885 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-533-306A-4

Query Match 4.0%; Score 205; DB 2; Length 885;
Best Local Similarity 20.8%; Pred. No. 1.1e-08;
Matches 105; Conservative 105; Mismatches 184; Indels 110; Gaps 20;

QY 40 TPYLQAKVLCQDEEVLHSP---RLTNSAMRAGHLLDLLKTR---GKNGAIAFLF--- 90
Db 430 TKALSARL-----EALAEKEELERTNKLKA-EMEDLVSSKDDVGKN--VHELEKSK 481

QY 91 -SLKFHPDVTYLTGLOPDVDFNSGLMETSCLTECLAGATGSLQEEINQKQKEVL 149
Db 482 RALETOMEENKTOLEEELEQASEDAKL----RLEVNMQALKGQFERDLQARDEQNE-- 535

QY 150 LRCQQLQELHGLAETRAEGLHLEADHSRMKREVSAPHFHVLRLKDEMLSLSLHYSNAL 209
Db 536 -EKRRQLQRLQHEYT-----ELEDNERALAAAKK--KLEGDLKDLQADSAI 584

QY 210 QEKELAAASRCRSLQEEYLLKQELQ-----RANMVSCELELQBSL 251
Db 585 KGREEAIKQLRKLAQAKMDFORELADARASRDEIFATAKENKKAKSLEADLMQLEDLA 644

QY 252 RTASDQSGDEELNRLKEE-----NEKLSLTFSLAEKDILEQSLDEARGSRQE 300
Db 645 AAEPRARQAOLEKEELAEELASSLSGRNALQDEKRRLEARIAQ---LEELEEEQGNMEA 701

QY 301 LVERIHSRLRA-----VAAERQREYWEKEQTLLQFQSKMACOL-----YR 344
Db 702 MSDRVKATQQAQLSNELATERSTAGKNESARQQ--LERQNKELRSLKHEMGAVKSKFK 760

QY 345 EKVNALQAQVCELOKQERDQAYSARDSAQREISQSLVEKDSLRQVFLTDQVCELRQLR 404
Db 761 STTAALEAKIAQLEEQVEQAREKQAA---TKSLQKQKKKLEILLQVEDERKMAEQYK 816

QY 405 QLOAEPPGVLKQEARTEPCPREKQRLVRMHAICPRDDSCSLVSTESLSDLSL-SATS 463
Db 817 EQAEKGNARVQLKRLQLEEAEEESQR-----INANRRKLQRELDTE 859

QY 464 SREL----VDSFRSSSPAPPSSQOS 483
Db 860 SNEAMGREVNALKSLKLRGPPPPQET 883

Search completed: June 13, 2002, 09:24:20
Job time: 153 sec